

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

UNIVERSIDAD POLITÉCNICA DE MADRID



TRABAJO FIN DE GRADO

**ANÁLISIS Y MEJORA DE UN SISTEMA DE
SEGMENTACIÓN TEMPORAL DE VÍDEO**

Eva Casado Arnáiz

Noviembre 2014



Grupo de Aplicación de Telecomunicaciones Visuales

Departamento de Señales, Sistemas y Radiocomunicaciones

**ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

Título:	Análisis y mejora de un sistema de segmentación temporal de vídeo.
Alumno:	Eva Casado Arnáiz
Tutor:	Faustino Ángel Sánchez García
Ponente:	Federico Álvarez García
Departamento:	Señales, Sistemas y Radiocomunicaciones

Miembros del tribunal calificador	
Presidente:	D. José Manuel Menéndez García
Vocal:	D. Federico Álvarez García
Secretario:	D. Juan Cires Martínez
Suplente:	D. Juan Mariano de Goyeneche Vázquez de Seyas

FECHA DE LECTURA:.... de de 2014.

CALIFICACIÓN:

RESUMEN

Este Trabajo de Fin de Grado consiste en el análisis, evaluación y mejora de un sistema de segmentación temporal de vídeo, embebido dentro de un programa de caracterización estética de vídeos que detecta cambios de plano a través de cortes, fundidos y encadenados.

En primer lugar se realiza un análisis del programa original empleando métricas que permitan cuantificar el rendimiento y detectar los principales problemas y su contexto, buscando patrones comunes que permitan enfocar las mejoras necesarias para solventar dichos problemas. A continuación, se proponen mejoras tanto funcionales como no funcionales, que son acometidas en la fase de diseño e implementación. Para las mejoras relacionadas con la segmentación temporal, se aplican técnicas existentes en el estado del arte y se adaptan a las necesidades del programa.

Finalmente, se evalúa el programa comprobando que las métricas que habían cuantificado los problemas han mejorado y detallando el posible trabajo futuro a realizar sobre el programa.

The purpose of this Bachelor Thesis is to cover the analysis, evaluation and improvement of a temporal video segmentation system. Such system is embedded within a video aesthetics characterization program which detects shot boundaries through cuts, fades and dissolves.

Firstly, an analysis of the original program is performed using metrics to measure the efficiency and to identify the key problems and their context, looking for common patterns to focus on the required improvements and thus, resolve these problems. Consequently, this project proposes functional and non-functional improvements, which are undertaken in the design phase. In order to carry out the improvements related to temporal video segmentation, some state of the art techniques are adapted to cover the needs of the program.

Finally, the program is evaluated verifying that the metrics which had been used to measure the problems have been improved, therefore providing several conclusions and detailing the possible future work to be done on the system.

PALABRAS CLAVE

OpenCV, procesamiento de vídeo, segmentación temporal, precisión, exhaustividad, detección de cortes, detección de fundidos, características estéticas.

ÍNDICE

1. INTRODUCCIÓN	6
1.1 CONTEXTO Y NECESIDAD	6
1.2 OBJETIVOS	7
1.3 METODOLOGÍA	7
1.4 ESTRUCTURA DE LA MEMORIA	8
2. ESTADO DEL ARTE Y HERRAMIENTAS DE TRABAJO.....	9
2.1 INTRODUCCIÓN	9
2.2 DETECCIÓN DE CORTES	9
2.3 COMPARATIVA OPENCV	11
2.4 OTRAS HERRAMIENTAS	11
3. ANÁLISIS	13
3.1 CONSTRUCCIÓN DEL MARCO DE TRABAJO.....	13
3.2 ANÁLISIS DE CAJA NEGRA	14
3.2.1 SITUACIÓN INICIAL	14
3.2.2 PROCESO DE ANÁLISIS	15
3.2.3 PRECISIÓN Y EXHAUSTIVIDAD	20
3.3 REQUISITOS FINALES.....	23
4. DISEÑO E IMPLEMENTACIÓN DE LAS MEJORAS Y EVALUACIÓN.....	25
4.1 MEJORAS NO FUNCIONALES	25
4.1.1 MIGRACIÓN DEL CÓDIGO	25
4.1.2 MODIFICACIÓN DEL TIPO DE VÍDEO DE ENTRADA	26
4.1.3 MODIFICACIÓN DEL FICHERO DE SALIDA	27
4.2 MEJORAS FUNCIONALES	28
4.2.1 MEJORA DEL MÓDULO DE GENERACIÓN DE DESCRIPTORES FINALES	32
4.2.2 FUNDIDOS EN CUALQUIER COLOR.....	33
4.2.3 MEJORA DE LA DETECCIÓN DE CORTES	35
4.3 CONCLUSIONES DE LAS MEJORAS	42
5. CONCLUSIONES Y TRABAJO FUTURO	45
6. BIBLIOGRAFÍA	47

ÍNDICE DE FIGURAS

Figura 3.1 Marcos de trabajo.....	13
Figura 3.2 Script de consola del caracterizador de vídeo original.....	15
Figura 3.3 Descriptores del script de consola del caracterizador de vídeo original	17
Figura 3.4 Reproducción lenta Casablanca	19
Figura 3.5 Proceso del análisis de caja negra	23
Figura 4.1 Comparativa entre el fichero vídeo.(formato) en XML y el fichero vídeo.(formato)_descriptor	28
Figura 4.2 Métodos de detección de los cortes de plano en el programa original.....	29
Figura 4.3 Fundido de entrada y fundido de salida	30
Figura 4.4 Detección de una imagen en un solo color, candidato a fundido.....	34
Figura 4.5 Comparativa de histogramas de complejidad de cámara	38
Figura 4.6 Media armónica de los falsos positivos de “The Amazing Spider-Man.....	39
Figura 4.7 Mejora de la detección de corte de plano.....	41

ÍNDICE DE TABLAS

Tabla 3.1 Cambios de plano del vídeo “Casablanca”.....	18
Tabla 3.2 Comparativa entre “Casablanca” y “Al final de la escapada”.....	19
Tabla 3.3 Precisión y exhaustividad de “Casablanca”	21
Tabla 3.4 Resumen de la métrica precisión y exhaustividad.....	21
Tabla 3.5 Resumen de la métrica precisión y exhaustividad para los cortes de plano, fundidos y encadenados.....	21
Tabla 3.6 Evolución de la velocidad media de planos (ASL)	22
Tabla 3.7 Resumen de las conclusiones de la fase de análisis	24
Tabla 4.1 Detección de valores específicos de "Casablanca".....	31
Tabla 4.2 Comparativa entre la métrica precisión y exhaustividad de “Gritos y Susurros”	35
Tabla 4.3 Comparativa entre la métrica precisión y exhaustividad de “The Amazing Spider-Man”	42
Tabla 4.4 Comparativa entre la métrica precisión y exhaustividad de “El Señor de los Anillos”	42
Tabla 4.5 Comparativa de la métrica precisión y exhaustividad entre el antes y el después de los cambios de plano totales.....	43
Tabla 4.6 Comparativa de la métrica precisión y exhaustividad entre el antes y después para los cortes de plano y los fundidos.....	43

1. INTRODUCCIÓN

1.1 CONTEXTO Y NECESIDAD

Este Trabajo de Fin de Grado surge de la necesidad de disponer de una herramienta de segmentación temporal de vídeo interoperable, robusta y flexible.

El montaje es una de las herramientas básicas de la comunicación audiovisual y, a tal efecto, la obtención de métricas que puedan cuantificar automáticamente y de forma fiel su aplicación a un vídeo sigue siendo una cuenta pendiente. La elección de la cantidad y tipo de cambios de plano nos da una idea de la intención que ha tenido el autor de un vídeo para transmitir determinadas sensaciones o mensajes. Dentro de un mismo vídeo se puede encontrar escenas, secuencias y planos. Se entiende que un plano es una tira continuada de vídeo creado a partir de la sucesión de fotogramas. Una secuencia es una tira continuada de planos y se denomina escena a las secuencias de planos que tienen características comunes tales como la luminosidad, movimiento de cámara, etc. Se puede observar que el plano es la base tanto de las secuencias como de las escenas, por tanto sin una correcta detección de los planos, no se consigue una correcta detección de las secuencias o de las escenas.

En este Trabajo de Fin de Grado se parte de una versión inicial [1] de un programa de ordenador que detecta diversos tipos de segmentación temporal y del que se analiza su funcionamiento y se evalúan sus prestaciones. Una vez analizado y evaluado el programa de detección, y por tanto conocido el porqué de los errores producidos, se procede al diseño de posibles mejoras, a su implementación en el programa y a la evaluación del efecto de las mismas.

La utilidad de la herramienta creada en este Trabajo puede ser extrapolada para diversos ámbitos, desde la detección de diferentes tipos de plano para la rotulación y grafismo automático, publicidad personalizada hasta la recomendación multimedia, en general todo aquello que implique resumen, indexación y recuperación de vídeo. Este Trabajo de Fin de Grado se va a centrar en la caracterización estética de películas, tal y como se podrá comprobar en la fase de análisis. Otra de las utilidades de esta herramienta es la de poder ser utilizada en tiempo real, lo cual permite analizar contenidos emitidos en directo o la caracterización estética durante el tiempo de visionado de una película.

1.2 OBJETIVOS

Como anteriormente se ha explicado, partiendo de un programa original, un caracterizador de vídeo¹ sobre el que se desea realiza un análisis de funcionalidad y una implementación de mejoras asociadas, se definen los objetivos que se proponen en este Trabajo:

1. Análisis del caracterizador de vídeo inicial para detectar si el procesado de los vídeos se realiza de manera correcta o si, por el contrario, hay errores o deficiencias en la detección de planos.
2. Aplicación adecuada de las metodologías de reingeniería de *software*, de tal forma que a partir del código fuente de partida se construya una nueva versión del programa de ordenador.
3. Corrección o mejora de características no funcionales que no se ajusten a los requisitos establecidos y que sean necesarios para la interoperabilidad con distintas aplicaciones u otros sistemas.
4. Corrección del código del programa inicial que permita eliminar los errores en la detección de cambios de plano y en la obtención de descriptores de los vídeos, e introducir nuevas mejoras que permitan incluir funcionalidades adicionales al programa, mejorando y aumentando las capacidades del caracterizador.

1.3 METODOLOGÍA

Para lograr los objetivos propuestos en este Trabajo de Fin de Grado, se ha dividido el trabajo en 6 etapas que han permitido llevar un mejor seguimiento del análisis y mejora del programa. Dichas etapas son:

1. Estudio y aprendizaje del entorno C++ y OpenCV: estudio y comprensión del lenguaje C++ y OpenCV 2.4.9 utilizando el programa Visual Studio 2012 Ultimate.
2. Estudio de la documentación y comprensión del funcionamiento del programa original.
3. Definición de los criterios de análisis: selección y definición de las características de los parámetros que se detectarán en los vídeos mediante el programa, fijando finalmente que dichos parámetros sean cortes, fundidos, encadenados y escenas.

¹ Se denomina caracterizador de vídeo a un programa que permite analizar vídeos obteniendo los cambios de plano realizados, así como otras características de bajo y alto nivel del vídeo.

4. Análisis del funcionamiento del programa para los criterios definidos: análisis de caja negra para evaluar el funcionamiento del programa, encontrar deficiencias y fijar prioridades de mejora.
5. Evaluación del programa y definición de estrategias: estudio del código en busca de los errores que se producen en la detección de los parámetros definidos y posterior evaluación. Selección de posibles estrategias para solventar o minimizar los errores.
6. Diseño y desarrollo de mejoras: especificación de las mejoras de los algoritmos de segmentación temporal e implementación de la nueva versión del programa de ordenador.
7. Pruebas y comparación de resultados: comparación del programa antes y después de implementar las mejoras mediante los indicadores establecidos en la etapa de análisis.

1.4 ESTRUCTURA DE LA MEMORIA

La memoria de este Trabajo de Fin de Grado se ha dividido en 5 capítulos principales entre los que se encuentran:

1. **Capítulo 1:** explicación de los motivos por los cuales se llevó a cabo este Trabajo de Fin de Grado y definición de los objetivos, así como la metodología seguida y la estructura de la memoria.
2. **Capítulo 2:** revisión del estado del arte particularizado a la segmentación temporal de vídeos, estudio del lenguaje de programación C++ y de la librería OpenCV y definición de las herramientas de trabajo que se han usado.
3. **Capítulo 3:** análisis del programa inicial detectando si los cambios de plano que se detectan en los vídeos son correctos o erróneos y obtención de las funcionalidades que es necesario mejorar.
4. **Capítulo 4:** diseño e implementación de las mejoras o soluciones definidas durante el proceso de análisis y evaluación del código tras dichas mejoras.
5. **Capítulo 5:** presentación de las conclusiones obtenidas y de las líneas futuras de trabajo relacionadas con el caracterizador de vídeo.
6. **Bibliografía:** enumeración de las referencias, tanto electrónicas como físicas, consultadas durante la realización del trabajo.

2. ESTADO DEL ARTE Y HERRAMIENTAS DE TRABAJO

2.1 INTRODUCCIÓN

El programa original, denominado “caracterizador de vídeo”, que se ha analizado en este Trabajo de Fin de Grado, está implementado en el lenguaje C++. Este lenguaje fue diseñado a mediados de los años 80 con el objetivo de extender el lenguaje C a la manipulación de objetos. Por lo tanto, el lenguaje C++ es un lenguaje híbrido ya que está orientado a objetos y surgió a raíz de un lenguaje que no estaba orientado a objetos.

Además, el programa original utiliza la biblioteca libre OpenCV, creada por Intel cerca del año 1999 [2] y la cual es muy utilizada hoy en día en sistemas de visión artificial, con aplicaciones directas en programas que tienen relación con la detección de movimiento o de control de procesos donde se hace necesario el reconocimiento de objetos. Por esta razón se convierte en una biblioteca ideal para utilizar en este trabajo debido a la necesidad de extraer características de vídeo de bajo nivel.

Un programa que permite trabajar con la biblioteca de OpenCV e implementar código en lenguaje C++ es el programa Visual Studio [3] El caracterizador de vídeo original del que parte este Trabajo estaba realizado en Visual Studio 2010 usando la biblioteca de OpenCV 2.2.0.

Existen diversas metodologías para la detección de cambios de plano, más concretamente para la detección de cortes de plano, por lo cual se hace necesario su estudio antes de profundizar en el caracterizador de vídeo original para comprender qué técnicas utiliza. Además, el uso de diferentes programas facilita el análisis y la mejora del caracterizador de vídeo, programas que se explican más adelante en este mismo capítulo.

2.2 DETECCIÓN DE CORTES

A la hora de realizar la detección de cortes de plano, se pueden distinguir dos procesos diferenciados, la primera es la cuantización para detectar los candidatos a corte de plano y la segunda es el filtrado para decidir cuáles de los candidatos son considerados como corte de plano.

En el proceso de cuantización, existen diversas técnicas para la detección de cortes que vienen siendo estudiadas desde hace varios años y que, actualmente, se siguen estudiando. Entre las técnicas tradicionales más efectivas se pueden encontrar [4] y [5]

- **Comparación por píxeles:** se basa en la comparación píxel a píxel entre dos cuadros consecutivos. Para detectar que dicha comparación es suficientemente importante como para ser considerada un corte de plano, se pueden calcular las diferencias absolutas o las diferencias entre las medias armónicas.
- **Comparación de bordes:** se basa en la comparación de la suma del número de bordes existentes entre dos cuadros consecutivos, indicando que existe corte si la diferencia es muy significativa.
- **Comparación del histograma de color:** se basa en la comparación de histogramas entre dos cuadros consecutivos. Cada histograma guarda la información del color de todos los píxeles de cada cuadro. Es una de las técnicas menos fiables ya que dos cuadros que sean completamente diferentes pueden tener histogramas muy similares. Además, no permite detectar cortes en vídeos en blanco y negro.
- **Comparación por movimiento:** se basa en comparar el movimiento entre dos cuadros consecutivos, indicando que se ha producido corte si dicho movimiento es muy abrupto entre los dos cuadros. Es una de las técnicas más actuales [6]

Recientemente están surgiendo nuevas técnicas de detección, aunque aún no están tan asentadas debido a su carácter más experimental y a su falta de validación en diferentes escenarios, por lo que simplemente se consideran como posibles ideas a la hora de implementar las mejoras, sin profundizar en su funcionamiento [7]

En el proceso de filtrado, se hace necesario contar con umbrales para establecer el valor a partir del cual se debe considerar que ha ocurrido un corte de plano. Dichos umbrales pueden ser de dos tipos:

- **Umbral fijo:** es un valor determinado que no varía en todo el vídeo. Normalmente es utilizado como valor máximo en las comparaciones a partir del cual se considera corte de plano.
- **Umbral variable:** se diferencia del anterior en que no es un valor predeterminado para todo el vídeo, sino que se recalcula conforme a la naturaleza de las escenas que existen en el vídeo. Es un valor mucho más preciso que el fijo y más fiable.

Una vez que se tienen seleccionados los cortes de plano, existen diversas métricas para cuantificar la correcta detección de los mismos. Una de las métricas más utilizadas es la de precisión y exhaustividad [8]

2.3 COMPARATIVA OPENCV

Debido a la cantidad de versiones existentes de la biblioteca OpenCV, se ha realizado una comparativa de la versión del código original implementado con OpenCV 2.2.0 y la última versión estable OpenCV 2.4.9.

La versión 2.2.0 [9] surge a finales del año 2010, por lo cual desde sus comienzos fue implementada en los programas que existían en aquellos años, como es el caso del programa Visual Studio 2010 donde fue desarrollada la primera versión del caracterizador de vídeo. Esta versión reorganiza las librerías e incorpora módulos nuevos, más específicos que “cxcore”, “cv”, “cvaux”, “highgui” y “ml” de las versiones anteriores, que facilitan la manipulación de vídeo.

A lo largo de los años, se ha intentado hacer una evolución de las librerías de OpenCV [9] para que cada vez se parezca más al lenguaje Python, lenguaje que también soporta orientación a objetos y cuya filosofía es la de favorecer un código legible. Además, se han ido introduciendo optimizaciones compatibles con OpenCL, lenguaje creado por Apple dedicado a entornos gráficos. Con esta intención surge en el año 2014 la versión OpenCV 2.4.9, año en el que todos los ordenadores que se venden en el mercado que tienen como sistema operativo Windows cuentan con la versión de Windows 8. Esta versión introduce mejoras en las optimizaciones de OpenCL, en visualizaciones 3D y métodos de captura de vídeo entre otras [9]

Por las razones anteriormente descritas, para este Trabajo se ha decidido utilizar la versión de OpenCV 2.4.9, alargando así la vida del caracterizador siendo más fácil en un futuro la actualización del código, y el programa Visual Studio 2012.

2.4 OTRAS HERRAMIENTAS

Además de la versión de OpenCV utilizada y del programa Visual Studio 2012, este Trabajo ha requerido de otras herramientas para poder llevarse a cabo. El programa ha sido desarrollado y ejecutado en un entorno de trabajo que contaba con el sistema operativo Windows 7 y también ha sido probado en Windows Server 2008 para comprobar que podía ejecutarse en un entorno servidor.

1. **Script de consola:** el programa cuenta con un *script* de consola que ha sido creado para poder ser editado y así indicar el vídeo, además de otros parámetros, que uno quiere que analice el programa y poder ser ejecutado mediante el propio *script* sin necesidad de abrir más programas. Este *script* puede ser editado con herramientas básicas como el Bloc de Notas de Windows, lo cual facilita al usuario la utilización del programa.
2. **aTube Catcher:** una de las acciones importantes que se ha llevado a cabo en este Trabajo es la de poder analizar diversos tipos de vídeo con diferentes formatos. Para poder descargar y dar el formato adecuado o deseado a los

vídeos, se ha utilizado el programa aTube Catcher, que permite descargar fácilmente vídeos de internet y transcodificarlos al formato deseado empleando una amplia batería de códecs y contenedores [10]

3. **VLC Media Player:** a la hora de analizar los vídeos y detectar si el programa ha detectado los cambios de plano de manera correcta, es necesario contar con un reproductor de vídeo que permita reproducir los vídeo a una velocidad lenta para así evitar los posibles fallos humanos visuales que se puedan cometer al analizar el vídeo. Este Trabajo se ha decantado por el reproductor VLC Media Player debido a su licencia libre, facilidad de manipulación y capacidad de reproducir vídeos en diferentes formatos que otros reproductores no son capaces de leer.
4. **Excel:** todos los datos que se recogen del análisis del vídeo se guardan en hojas de cálculo del programa Excel del paquete de Microsoft Office que se encuentra en el sistema operativo Windows en el que está realizado este Trabajo. Las tablas que se realizan en este programa con los datos obtenidos permiten identificar con mayor facilidad las propiedades de cada cambio de plano y definir por qué se ha producido cada uno de ellos. Además, el poder crear diferentes hojas de cálculo dentro del mismo libro y la facilidad de uso son las razones de la utilización de este programa para este Trabajo.

3. ANÁLISIS

Esta parte de la memoria recoge el proceso de análisis que se ha seguido en este trabajo. Se divide en la construcción del marco de trabajo, el análisis de caja negra realizado sobre el caracterizador de vídeo original y las métricas que se calculan de dicho análisis, seguido de los requisitos finales (a modo de conclusión del apartado) que se diseñan como mejoras para el programa.

3.1 CONSTRUCCIÓN DEL MARCO DE TRABAJO

Mientras que en el capítulo 2 se han detallado las herramientas de trabajo que han sido utilizadas, en este capítulo se explica la manera en que han sido usadas en un entorno de trabajo diario para desarrollar la fase de análisis. Dividiendo el trabajo en dos fases, análisis e implementación de las mejoras, se dispone de dos entornos más diferenciados:

- **Análisis:** en esta fase del Trabajo se usa principalmente el *script* de consola del programa original mediante el cual se analizan los vídeos y películas. Lo primero de todo será seleccionar y descargar diferentes tipos de vídeos mediante el programa aTube Catcher y analizarlos con el programa original en el que se basa este Trabajo para extraer los cambios de plano que se han detectado en los vídeos seleccionados. Una vez extraídos estos datos, se guardan en una tabla de una hoja de cálculo del programa Excel del paquete de Microsoft Office. Mediante el VLC Media Player se reproducen los vídeos analizados por el caracterizador y se clasifican los cambios de plano detectados por el programa guardando estos nuevos datos en la misma hoja de cálculo. Gracias a los resultados obtenidos se recopilan los errores y los patrones comunes con que aparecen los mismos y se diseñan las mejoras.

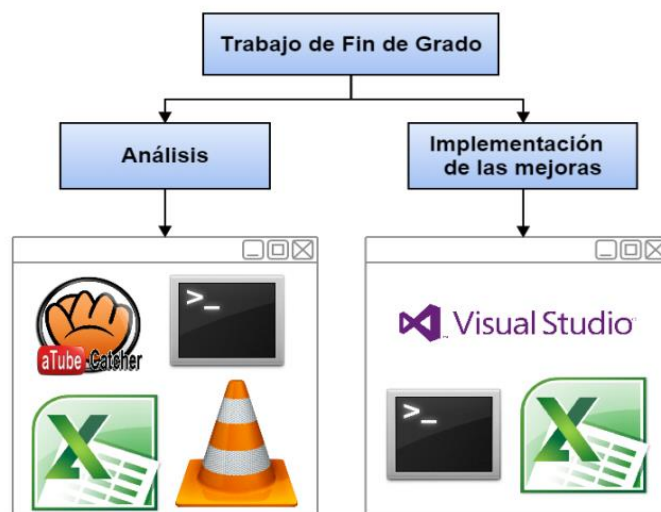


Figura 3.1 Marcos de trabajo

- **Diseño e implementación de las mejoras:** en esta fase del Trabajo, se introduce el programa Visual Studio 2012, ya que es necesario modificar el código para imprimir por la pantalla de la consola de Windows diferentes datos que faciliten la descripción en mayor detalle de cada cambio de plano. Estos datos que se imprimen por pantalla se guardan, junto con los datos recopilados en la primera fase, en la misma hoja de cálculo de Excel. Una vez implementadas las mejoras, los datos que se obtienen de dichas mejoras se guardan con los anteriores.

3.2 ANÁLISIS DE CAJA NEGRA

Esta parte del tercer capítulo supone una de las partes principales del documento. En ella, se detallan los primeros pasos realizados en el Trabajo, donde se obtienen los primeros datos en los que se basará el resto de estudios e implementaciones realizadas sobre el programa original.

3.2.1 SITUACIÓN INICIAL

Como ya se ha mencionado anteriormente en el documento, se parte de un caracterizador automático de vídeo que extrae las principales propiedades de bajo nivel de los vídeos, tales como descriptores de imagen, ritmo narrativo o descriptores de movimiento. Entre las propiedades que detecta se encuentra la detección de los cambios de plano, propiedad en la que se basa principalmente este Trabajo y en la cual se ha profundizado en mayor medida.

Este proyecto ha tratado analizar el mayor número de efectos de postproducción y grabación posibles. A lo largo de los años estos métodos han ido cambiando debido a que la demanda y el gusto de los espectadores ha ido evolucionando y a que han aparecido nuevos géneros y estéticas. En este Trabajo se busca analizar la mayor diversidad posible de vídeos que puedan ser una referencia de los errores que se encuentran en esta fase de análisis. Por estos motivos se seleccionan en total diez vídeos cortos, correspondientes a un fragmento de una película cuyo título se va a utilizar para su denominación, y una película completa. Entre los vídeos cortos se seleccionan cinco vídeos en blanco y negro y cinco en color, además de contar tanto con películas lentas, pertenecientes a géneros como el drama o la comedia, como películas que cuentan con mayor movimiento y ritmo visual, como las películas de acción. Finalmente los vídeos que se han seleccionado son:

- **Vídeos en blanco y negro:** “Al final de la escapada” (1960), “El Apartamento” (1960), “Casablanca” (1942), “Enemigo público” (1930) y “Con faldas y a lo loco” (1959). Películas que han salido al público entre los años 1930 y 1960, la mayoría de ellas pertenecientes al género del drama, romance o la comedia y una de ellas al cine negro.

- **Vídeos en color:** “Apocalipsis Now” (1979), “The Amazing Spider-Man” (2012), “El Señor de los Anillos” (2001), “Titanic” (1997) y “Gritos y Susurros” (1972). Películas entre los años 1972 y 2012 que abarcan géneros como el drama, romance, fantasía o la ciencia ficción.

La película completa que se ha analizado ha sido “Hannah y sus hermanas”, adecuada debido a que la propia película incluye escenas en las que se emite una película en blanco y negro y a que presenta diferentes tipos de transiciones y escenas diferentes. Además, gracias a esta película se puede comprobar el rendimiento del caracterizador con vídeos de larga duración y realizar pruebas de regresión una vez introducidas las mejoras.

Todas estas películas se encuentran en formato .avi debido a que el caracterizador inicial del que partía este Trabajo sólo era capaz de analizar vídeos en este tipo de formato. Es por ello que aquellos vídeos de los que no había disponibilidad, se bajaron y transcodificaron mediante el programa aTube Catcher.

3.2.2 PROCESO DE ANÁLISIS

Una vez seleccionadas las películas, se desarrolla la parte fundamental de este Trabajo. En primer lugar, se ejecuta el programa inicial mediante su script de consola, en la cual se puede encontrar información de gran utilidad:

```

C:\Windows\system32\cmd.exe
Hora: 10:07:11,48 → Hora a la que se realiza el análisis
unFrames:2889 → Número de cuadros y cuadros por segundo del video
fps:29.970030
casablanca.avi → Nombre del video.(formato)

Iniciamos flujo con los siguientes valores de deteccion de movimiento:
Iteraciones: 20
Apertura: 10
Error: 0.001000

CORTE 80
Analizado el 2.80 por ciento de la pelicula

CORTE 298
Analizado el 10.35 por ciento de la pelicula

CORTE 1076
Analizado el 37.28 por ciento de la pelicula

CORTE 1146
Analizado el 39.70 por ciento de la pelicula

Corte por transicion
  
```

Figura 3.2 Script de consola del caracterizador de vídeo original

Se puede observar que en la consola de Windows se aprecian diversos datos que van a ser de utilidad. Es más, en este documento se verá más adelante que se pueden modificar los datos que se quiere que se impriman por pantalla, lo cual facilitará la búsqueda de aquellas características del vídeo que el caracterizador analice y que interese para el desarrollo del trabajo. Tal y como se observa en la Figura 3.2 los primeros datos que se extraen del vídeo son sencillos, tales como la hora de ejecución

del programa, número de cuadros y cuadros por segundo del vídeo, nombre del vídeo y su formato y los valores iniciales preestablecidos de detección del movimiento.

Además de estos datos, se observan los primeros cambios de plano que ha detectado el caracterizador. Estos cambios de plano se imprimen en la consola seguidos del porcentaje que lleva analizado del vídeo. Se pueden diferenciar en:

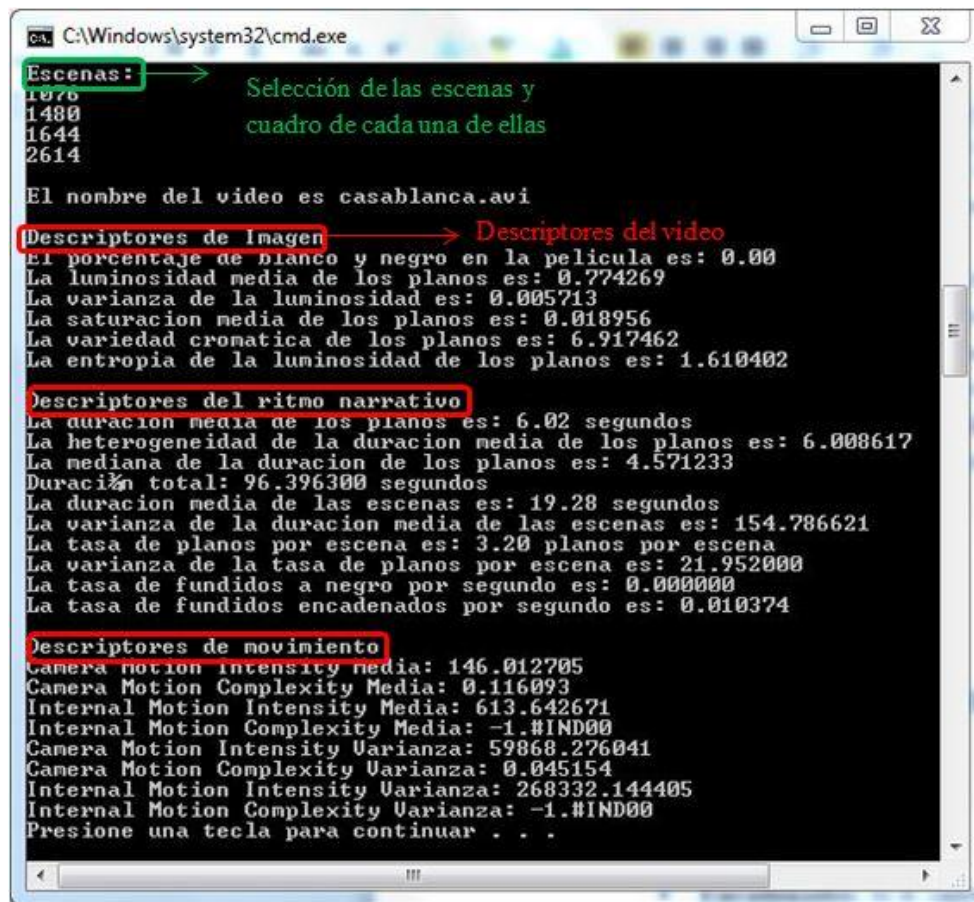
- **Corte de plano:** es el cambio de plano que se produce al cambiar abruptamente de un plano. Es el cambio de plano más común y en la consola aparecerá con el nombre de “CORTE” precedido por el número de cuadro en el que se produce el corte.
- **Fundidos:** es el cambio de plano que se produce al cambiar de un plano a otro monocolor con una transición más o menos pronunciada. Normalmente se produce en los comienzos y finales de las películas, pero esto no implica que esta técnica no se use durante el desarrollo normal de la película. En la consola aparece como “*fade in*” o “*fade out*”, indicando si el plano monocolor es anterior o posterior, respectivamente, al plano normal.
- **Encadenados:** es el cambio de plano que se produce al cambiar de un plano a otro con una transición, esto es, un plano se funde en otro plano a mayor o menor velocidad. Son poco frecuentes en películas contemporáneas, pero muy abundantes en películas antiguas, y en la consola aparecen como “corte por transición”.
- **Escenas:** son la sucesión de planos, puede contener cuales quiera de los cambios de planos antes descritos, que tienen características en común. Está delimitado principalmente por los cambios de plano. En la consola se indicarán al final del análisis del vídeo y aparecerán indicadas por el número de cuadro final de la misma, tal y como se puede observar en la *Figura 3.3*.

Cuando se termina de analizar un vídeo, se imprimen por pantalla algunos descriptores del mismo. Aunque en principio pueda parecer que carezcan de utilidad para esta fase de análisis del Trabajo, en las fases posteriores se utilizarán para diseñar e implementar las mejoras necesarias.

Además de imprimir estos datos por la consola de Windows, el caracterizador genera hasta cuatro archivos de texto que contienen:

- **Vídeo.(formato):** contiene los valores ordenados de los descriptores de imagen, de ritmo narrativo y de movimiento.
- **Vídeo.(formato)_cortes:** contiene todos los cambios de plano que el caracterizador ha detectado, indicando el número de cuadro y el

momento en formato de tiempo en el que se han producido dentro de la duración del vídeo.



```
C:\Windows\system32\cmd.exe
Escenas: → Selección de las escenas y
          cuadro de cada una de ellas
1076
1480
1644
2614

El nombre del video es casablanca.avi

Descriptores de Imagen → Descriptores del video
El porcentaje de blanco y negro en la película es: 0.00
La luminosidad media de los planos es: 0.774269
La varianza de la luminosidad es: 0.005713
La saturación media de los planos es: 0.018956
La variedad cromática de los planos es: 6.917462
La entropía de la luminosidad de los planos es: 1.610402

Descriptores del ritmo narrativo
La duración media de los planos es: 6.02 segundos
La heterogeneidad de la duración media de los planos es: 6.008617
La mediana de la duración de los planos es: 4.571233
Duración total: 96.396300 segundos
La duración media de las escenas es: 19.28 segundos
La varianza de la duración media de las escenas es: 154.786621
La tasa de planos por escena es: 3.20 planos por escena
La varianza de la tasa de planos por escena es: 21.952000
La tasa de fundidos a negro por segundo es: 0.000000
La tasa de fundidos encadenados por segundo es: 0.010374

Descriptores de movimiento
Camera Motion Intensity media: 146.012705
Camera Motion Complexity Media: 0.116093
Internal Motion Intensity Media: 613.642671
Internal Motion Complexity Media: -1.#IND00
Camera Motion Intensity Varianza: 59868.276041
Camera Motion Complexity Varianza: 0.045154
Internal Motion Intensity Varianza: 268332.144405
Internal Motion Complexity Varianza: -1.#IND00
Presione una tecla para continuar . . .
```

Figura 3.3 Descriptores del script de consola del caracterizador de vídeo original

- **Vídeo.(formato)_descriptor:** contiene los valores ordenados de los descriptores de imagen, de ritmo narrativo y de movimiento, pero a diferencia del primer archivo, en este está claramente especificado a qué corresponden cada uno de los valores. Es un archivo similar a lo que se imprime por la consola al final del análisis.
- **Vídeo.(formato)_escenas:** Al igual que ocurre con el resto de cambios de plano, indica el cuadro y el momento en el que se han producido.

Una vez visto qué elementos aparecen durante el proceso de análisis de un vídeo, se procede a explicar el análisis de todos los vídeos seleccionados entendiendo que ya se tienen los archivos de texto, generados por el caracterizador, de todos ellos.

Se crea una tabla en una hoja de cálculo del programa Excel para cada uno de los vídeos analizados. Cada tabla contiene los cambios de plano que el caracterizador ha detectado, indicando el número del cuadro en el cual se ha producido el cambio de plano y el momento temporal en el que se produce. Es decir, en cada una de estas tablas se van a recoger los datos que se obtienen de los ficheros vídeo.(formato)_cortes para así poder analizar, uno por uno, los cambios de plano que el caracterizador ha detectado.

También se indica el número de cuadros por segundo que contiene el vídeo, lo cual permite obtener el instante temporal en el que se produce cada cambio de plano. En la *Tabla 3.1* se puede observar un ejemplo de lo que sería la tabla inicial que se crea. Al igual que se observa este ejemplo para el vídeo de “Casablanca”, se crean tablas análogas para cada uno de los 10 vídeos analizados.

Una vez que se crean estas tablas, se analiza cada una de las películas para comprobar que los cortes de plano detectados por el caracterizador son correctos; para ello, se reproduce con el VLC Media Player cada uno de los vídeos y se comprueba manualmente que se produce un corte de plano en el tiempo que se indica; además se comprueba si existen otros cambios de plano además de los detectados. Como no todos los vídeos tienen la misma velocidad de reproducción ni las mismas transiciones, se utiliza la reproducción lenta para poder observar con mayor detenimiento los cambios de plano de los vídeos.

Casablanca	Cambios de plano detectados
29,97 F/s	1. 80 ---> 0:0:2
	2. 298 ---> 0:0:9
	3. 1076 ---> 0:0:35
	4. 1146 ---> 0:0:38
	5. 1241 ---> 0:0:41
	6. 1301 ---> 0:0:43
	7. 1480 ---> 0:0:49
	8. 1584 ---> 0:0:52
	9. 1644 ---> 0:0:54
	10. 1811 ---> 0:1:0
	11. 2003 ---> 0:1:6
	12. 2172 ---> 0:1:12
	13. 2614 ---> 0:1:27
	14. 2696 ---> 0:1:29
	15. 2833 ---> 0:1:34

Tabla 3.1 Cambios de plano del vídeo “Casablanca”

La opción de reproducción lenta se encuentra dentro de la barra de tareas tal y como se puede observar en la Figura 3.4. Gracias a que en las tablas generadas se ha indicado la velocidad de cada uno de los vídeos, se puede saber si el número de cuadro indicado de cada uno de los cambios de plano tiene mayor o menor error ya que el ojo humano no es capaz de detectar el cambio de plano con la suficiente precisión, tan solo hacerse una idea del segundo en el que se ha producido.

Tal y como se ha dicho anteriormente, se analiza cada uno de los cambios de plano detectados. En las tablas anteriormente creadas, se indica la situación de cada uno de estos cambios de plano, no indicando nada en especial si son correctos o explicando si es incorrecto o si falta algún otro que no ha sido identificado. A continuación se

expone, a modo de ejemplo, un par de tablas, *Tabla 3.2*, para ver el contraste entre dos vídeos en los cuales se han detectado cambios de planos incorrectos o no identificados.

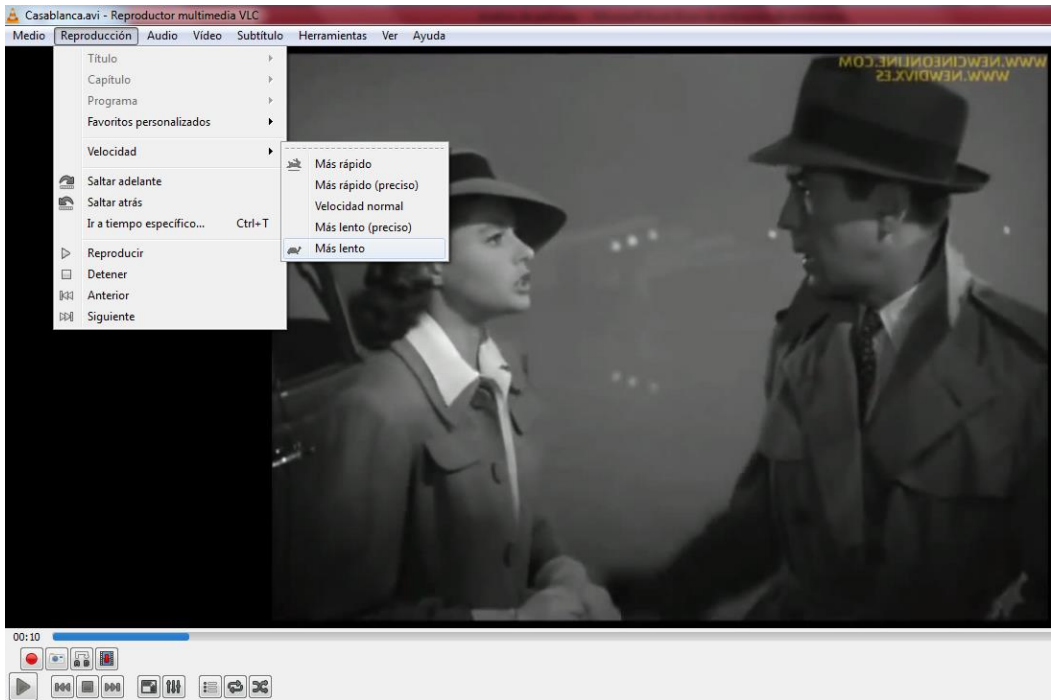


Figura 3.4 Reproducción lenta Casablanca

Casablanca	Cambios de plano detectados	Comentarios	Al final de la escapada	Cambios de plano detectados	Comentarios
29,97 F/s	1. 80 ---> 0:0:2		30 F/s	1. 149 ---> 0:0:4	
	2. 298 ---> 0:0:9			2. 442 ---> 0:0:14	
	3. 1076 ---> 0:0:35			3. 597 ---> 0:0:19	
	4. 1146 ---> 0:0:38			4. 1337 ---> 0:0:44	
	5. 1241 ---> 0:0:41	Zoom de cámara, falso positivo		5. 1631 ---> 0:0:54	
	6. 1301 ---> 0:0:43			6. 1802 ---> 0:1:0	
	7. 1480 ---> 0:0:49			7. 1975 ---> 0:1:5	Falta corte antes del 1:8
	8. 1584 ---> 0:0:52			8. 2062 ---> 0:1:8	
	9. 1644 ---> 0:0:54			9. 2110 ---> 0:1:10	
	10. 1811 ---> 0:1:0			10. 2156 ---> 0:1:11	Falta corte sobre el 1:12
	11. 2003 ---> 0:1:6			11. 2543 ---> 0:1:24	
	12. 2172 ---> 0:1:12				
	13. 2614 ---> 0:1:27				
	14. 2696 ---> 0:1:29				
	15. 2833 ---> 0:1:34				

Tabla 3.2 Comparativa entre “Casablanca” y “Al final de la escapada”

Como ya se ha expuesto anteriormente en el documento, este análisis se ha realizado con todos los vídeos seleccionados, teniendo en este punto del Trabajo once tablas como las de la *Tabla 3.2*, en las cuales se ven cada uno de los cambios de plano analizados, y puestas todas ellas en la misma hoja de cálculo. Por razones de espacio, no se incluyen en esta memoria.

3.2.3 PRECISIÓN Y EXHAUSTIVIDAD

La precisión y exhaustividad [11] es una métrica empleada en la medida del rendimiento de los sistemas de búsqueda y recuperación de información y reconocimiento de patrones. La precisión [12] por definición, es el ratio entre el número de documentos relevantes recuperados y el número de documentos recuperados, mientras que la exhaustividad [13] es el ratio entre el número de documentos relevantes recuperados y el número de documentos relevantes. Tal y como se ha comentado en el apartado del Estado del Arte, es una de las métricas más utilizadas para la prueba y validación de sistemas de segmentación temporal de vídeo.

En todo el Trabajo, contar con ratios que permitan acotar los resultados para poder clasificarlos en positivos o negativos, permite evaluar de manera más precisa los resultados obtenidos. Se ha considerado que en el análisis de los vídeos realizado, es de gran utilidad aplicar estos dos ratios para tener un valor acotado de su exactitud, poder realizar comparativas entre el análisis de diferentes vídeos y evaluar los resultados.

Para poder calcular estos ratios, se hace necesario definir esta métrica de acuerdo a los objetivos de este Trabajo y así relacionar cada parámetro definido con los datos que se encuentran en las tablas de la hoja de Excel anteriormente comentada:

$$\text{Precisión} = \frac{\text{cambios de plano detectados correctos}}{\text{cambios de plano detectados}}$$

$$\text{Exhaustividad} = \frac{\text{cambios de plano detectados correctos}}{\text{cambios de plano correctos}}$$

Se puede observar que lo que en la definición se menciona como documentos relevantes recuperados pasa a ser cambios de plano detectados correctos ya que son aquellos datos obtenidos que son relevantes para el análisis. Del mismo modo se identifica el resto de parámetros: los documentos recuperados van a ser los cambios de plano detectados y los documentos relevantes pasan a ser cambios de plano correctos, ya que son los datos que realmente importan.

Una vez definidos los ratios de la manera adecuada, se realiza un recuento de los parámetros necesarios en cada uno de los vídeos y se calculan los ratios precisión y exhaustividad para cada uno de ellos. Completando la hoja de cálculo anteriormente mencionada, se crean tablas adicionales que contienen los resultados, como la *Tabla 3.3*, que se muestra a continuación:

Casablanca	
Total cambios de plano detectados	15
Total cambios de plano detectados correctos	14
Total cambios de plano correctos	14
Precisión	93,33%
Exhaustividad	100,00%

Tabla 3.3 Precisión y exhaustividad de “Casablanca”

Una vez realizadas las tablas para cada uno de los vídeos, se crean unas tablas a modo resumen, *Tabla 3.4* y *Tablas 3.5*, que contengan todos los resultados de precisión y exhaustividad de los vídeos.

Película	C/B&W	Resolución	Duración	Precisión	Exhaustividad
Hannah y sus hermanas	Color	520x272	1:42:21	94,68%	98,42%
Al final de la escapada	Blanco y negro	480x360	0:01:25	100,00%	84,62%
El apartamento	Blanco y negro	480x360	0:04:29	100,00%	28,57%
Casablanca	Blanco y negro	960x720	0:01:36	93,33%	100,00%
Enemigo público	Blanco y negro	496x360	0:02:21	100,00%	90,48%
Con faldas y a lo loco	Blanco y negro	352x288	0:01:41	100,00%	93,75%
Apocalipsis Now	Color	480x360	0:03:08	93,94%	86,11%
The Amazing Spider-Man	Color	1280x540	0:02:05	66,00%	91,67%
El Señor de los Anillos	Color	1280x714	0:02:46	95,24%	83,33%
Titanic	Color	1920x1080	0:02:01	100,00%	100,00%
Gritos y Susurros	Color	320x240	0:10:03	97,50%	84,78%

Tabla 3.4 Resumen de la métrica precisión y exhaustividad

Película	Cortes		Fundidos		Encadenados	
	Precisión	Exhaustiv.	Precisión	Exhaustiv.	Precisión	Exhaustiv.
Hannah y sus hermanas						
Al final de la escapada	100,00%	84,62%	-	-	-	-
El apartamento	100,00%	25,00%	-	-	100,00%	50,00%
Casablanca	100,00%	100,00%	-	-	0,00%	-
Enemigo público	100,00%	100,00%	-	-	-	0,00%
Con faldas y a lo loco	100,00%	93,75%	-	-	-	-
Apocalipsis Now	96,88%	86,11%	-	-	0,00%	-
The Amazing Spider-Man	66,00%	91,67%	-	-	-	-
El Señor de los Anillos	95,18%	82,29%	-	-	0,00%	-
Titanic	100,00%	100,00%	100,00%	100,00%	-	-
Gritos y Susurros	100,00%	95,12%	0,00%	0,00%	-	0,00%

Tabla 3.5 Resumen de la métrica precisión y exhaustividad para los cortes de plano, fundidos y encadenados

Para comprender las tablas, el valor de “-“significa que no existen datos para este cambio de plano, es decir que no se han detectado ningún cambio de plano de este tipo y que además no existen en este vídeo.

De la *Tabla 3.4* y *Tabla 3.5* se puede observar que hay varios vídeos que no cuentan con un ratio del 100% de precisión o exhaustividad, luego no son bien analizados por el caracterizador. Otro dato que se puede observar es que se tiene un vídeo que cuenta con el 100% de acierto tanto en precisión y exhaustividad. Ese vídeo pertenece a “Titanic” y el motivo por el cual tiene un análisis perfecto se debe tanto a su alta resolución como a las características estéticas de la película, con un ritmo de montaje moderado y un movimiento de cámara homogéneo y bastante sencillo. Por tanto se puede suponer que cuanto mayor calidad de vídeo mejor análisis va a realizar de los vídeos y cuanto peor, mayor número de fallos se va a obtener, como es el caso del vídeo “El apartamento”. La calidad de vídeo principalmente viene dada por la resolución y la calidad de codificación.

También se puede observar que todos los vídeos en blanco y negro cuentan con un valor del 100% en alguno de sus ratios, esto significa que el número de fallos que hay en el análisis de este tipo de vídeos va a ser inferior al de las películas en color. Además de la calidad de cada vídeo, un motivo por el cual esta métrica va a ser más positiva es debido a que estas películas datan de años anteriores a las de color, años en los cuales la velocidad media de planos y la velocidad de montaje eran más lentas. Se puede observar en un artículo de David Bordwell, de la Universidad de California [14] a través del análisis del ASL (*Average Shot Length*) cómo se ha ido aumentando la velocidad media entre planos en las películas a lo largo de los años, *Tabla 3.6*:

Años	Velocidad media común	Velocidad media en escenas de alto ASL
1930-1960	8-11 segundos	
1960-1970	6-8 segundos	3-4 segundos
1970-1980	5-8 segundos	4-5 segundos
1980-1990	5-7 segundos	4-5 segundos
1990-actualmente	<5 segundos	<2 segundos

Tabla 3.6 Evolución de la velocidad media de planos (ASL)

Es importante destacar que en géneros como la comedia o el drama y escenas como conversaciones, tradicionalmente con velocidad media de planos grandes, el montaje se ha vuelto en los últimos años mucho más rápido, estando comúnmente por debajo de los 4 segundos. Es decir, ya no sólo las películas de acción son rápidas.

Otro dato a tener en cuenta en la tabla de resultados es que, cuanto mayor duración tenga el vídeo, menos va a afectar un cambio de plano detectado de manera errónea o un cambio de plano no detectado, ya que el número de cambios de plano detectados va a ser muy superior a estos dos. Por esa razón, aunque en la película completa, “Hannah y sus hermanas”, se puede observar un número elevado de cambios de plano erróneos, sus valores de precisión y exhaustividad serán superiores a otros

vídeos como por ejemplo “Apocalipsis Now”, que presentan escenas especialmente difíciles para la detección de cortes.

Así pues, a la hora de diseñar las mejoras, este Trabajo se centra en aquellos vídeos con los que el caracterizador presenta un peor rendimiento, analizando el porqué de ese valor de ratio, buscando características comunes y proponiendo posibles soluciones.

3.3 REQUISITOS FINALES

Como ya se ha detallado en el documento, se ha realizado un proceso de análisis denominado análisis de caja negra debido a que no se ha modificado ni examinado aún el código original del caracterizador. De este análisis, cuyo proceso se resume en la *Figura 3.5*, se obtiene una serie de conclusiones finales.



Figura 3.5 Proceso del análisis de caja negra

En el apartado 3.2.3 ya se han detallado algunas de las conclusiones finales, en especial las que tienen que ver con el tipo de vídeo analizado, que se han extraído de este análisis de caja negra. A parte de observar el porqué de los valores de los ratios obtenidos, este apartado se centra en exponer las soluciones que se propusieron para la siguiente fase del Trabajo.

Una de las conclusiones que se ha obtenido es el hecho de que, debido a la versión del código, se consideró necesaria la migración del código a una versión de OpenCV más actualizada. Dicha versión, como se ha visto en el capítulo 2, fue la de OpenCV 2.4.9, debido a su actualidad y a las posibilidades de manipulación de imágenes de las que dispone frente a otras versiones de esta librería.

Las demás conclusiones obtenidas detalladas en el anterior apartado, hicieron necesario diseñar soluciones para mejorar los ratios obtenidos gracias a la métrica precisión y exhaustividad. Estas soluciones que se diseñaron, se dividieron en dos grupos claramente diferenciados: mejoras no funcionales y mejoras funcionales.

1. **Mejoras no funcionales:** se refieren a aquellas mejoras que no modifican en primera instancia la funcionalidad del código, es decir, que no tratan de mejorar el código como tal para aumentar el valor obtenido de los resultados de las métricas, sino que van a tratar de añadir propiedades complementarias al programa que aumentarán su atractivo.
2. **Mejoras funcionales:** se refieren a aquellas mejoras que surgen como solución a los problemas que el caracterizador de vídeos original tiene para diferenciar y detectar algunos de los cambios de plano obtenidos en el análisis de los vídeos. Estas mejoras, por tanto, requieren de la corrección del código original y de la posible implementación de nuevos métodos.

En la *Tabla 3.7* se puede observar un resumen de los problemas encontrados en el caracterizador en la fase de análisis y su correspondientes posibles mejoras, tanto no funcionales como funcionales, que se le aplicarán al caracterizador de vídeo en la siguiente fase del Trabajo.

Problema encontrado	Solución	Mejora no funcional	Mejora funcional
No se puede modificar el código para añadir mejoras	Actualización de librerías de OpenCV y refactorización del código	Migración del código	
El caracterizador de vídeo únicamente reconoce vídeos en formato .avi	Modificación del programa para que admita otros formatos de vídeo	Modificación del tipo de vídeo de entrada	
El caracterizador de vídeo genera todos los archivos de salida en formato .txt	Modificación de uno de los ficheros de salida .txt a .xml	Modificación del fichero de salida	
El caracterizador de vídeo se bloquea cuando intenta analizar vídeos con un único plano	Modificación del código en el módulo de generación de los descriptores finales		Mejora del módulo de generación de descriptores finales
El caracterizador de vídeo sólo se reconoce fundidos en negro	Modificación del algoritmo de detección de fundidos		Fundidos en cualquier color
Métrica precisión y exhaustividad pobre en algunos vídeos	Modificación del algoritmo de detección de cortes		Mejora de la detección de cortes

Tabla 3.7 Resumen de las conclusiones de la fase de análisis

4. DISEÑO E IMPLEMENTACIÓN DE LAS MEJORAS Y EVALUACIÓN

Esta parte de la memoria está estructurada en función de las mejoras que se realizan en el caracterizador de vídeo encontrando, por un lado, las mejoras no funcionales y la explicación de cada una de ellas, y por otro lado las funcionales y su explicación. La evaluación de cada mejora está incluida en cada apartado correspondiente.

4.1 MEJORAS NO FUNCIONALES

Una vez desarrollada la fase de análisis detallada en el capítulo 3 de este documento, se diseña una serie de mejoras o soluciones que se implementan en el caracterizador de vídeo original. En el último apartado de dicho capítulo se explicaba de manera breve las razones por las cuales se hace necesario abordar dichas mejoras.

En este apartado, se explican las necesidades que llevaron al diseño de mejoras no funcionales que le dan un valor añadido al programa inicial.

4.1.1 MIGRACIÓN DEL CÓDIGO

La migración del código original a la versión 2.4.9 surge de la necesidad de modificar la implementación de dicho código para añadirle mejoras y solucionar los errores que introducía en el análisis de los vídeos. Se escoge esta versión de OpenCV por las razones que se comentan en la sección 2.2 de este documento.

Una vez que se ha completado la migración, se puede modificar el código e introducir trazas² en el mismo para poder analizar en mayor profundidad los cambios de plano, observando el comportamiento de los algoritmos y cruzándolo con los resultados del análisis realizado.

Esta mejora no funcional, ocupó una gran parte del tiempo del Trabajo debido al estudio previo del lenguaje C++ y de las versiones 2.2.0 y 2.4.9 de OpenCV que fueron necesarias para la comprensión adecuada del código del caracterizador de vídeo original y así realizar la migración sin cometer errores que impidiesen compilar el caracterizador mediante el programa Visual Studio 2012.

Una vez realizada la migración, se comprobó que funcionaba igual que la versión anterior y así se posibilitaba la implantación del resto de las mejoras funcionales y no funcionales.

² Se escriben líneas de código que permitan imprimir por la consola de Windows los valores necesarios para definir los diversos cambios de plano.

4.1.2 MODIFICACIÓN DEL TIPO DE VÍDEO DE ENTRADA

Observando el caracterizador de vídeo, se pudo ver que el único tipo de vídeos que era capaz de analizar correspondía al formato contenedor .avi. Este era el formato más utilizado para almacenar vídeo digital [15] hasta la aparición de códecs preparados para el procesamiento de vídeo en alta definición. Por este motivo, es de suponer que en la primera versión del programa se pensase en este tipo de formato de vídeo para poder ser analizado. Sin embargo, existen otros tipos de formatos bastante utilizados en la actualidad debido a la finalidad que el usuario quiera dar a esos vídeos.

Aparte del tipo de formato de vídeo que el caracterizador de vídeo original era capaz de detectar, es necesario explicar que cada tipo de formato puede ser codificado con diferentes códecs. Gracias a la extensión ffmpeg que utiliza OpenCV para transcodificar los vídeos, el caracterizador de vídeo es capaz de analizar los vídeos, cuyo formato sea reconocido por el propio programa, que estén codificados en cada uno de los códecs que aparecen en la referencia [16]

Como ya se ha explicado, existen diversos formatos dependiendo de quién los haya producido o de su finalidad, ya que cada formato cuenta con unas características que permiten ajustar la calidad y compresión del vídeo dependiendo del fin que se le espere dar.

Por este motivo, se plantea la necesidad de realizar pruebas con vídeos con diferentes tipos de formato e implementar esta propiedad en el caracterizador de vídeo como una mejora no funcional.

Una vez implementado el código para que reconozca, a priori, todo tipo de formatos de vídeo, se analizó su capacidad de detectar al menos los siguientes formatos de vídeo:

- **AVI:** (Audio Video Interleaved) formato reconocido por el caracterizador de vídeo original y anteriormente mencionado. Formato estándar para almacenar vídeo digital.
- **MP4:** formato cuyo códec de compresión puede ser tanto MPEG-4, capa 2 como H.264 (AVC) está orientado para vídeos en la web.
- **MOV:** formato desarrollado por Apple que utiliza un códec propio y que también es muy utilizado para la publicación de vídeos en internet debido a su relación calidad-peso.
- **WMV:** formato desarrollado por Microsoft que utiliza el códec MPEG-4 para la compresión de vídeo y que, aparte de ser utilizado para publicar vídeos en internet debido a su relación calidad-peso, admite *streaming*.

- **RM:** formato desarrollado por Real Networks que se puede encontrar en vídeos publicados en internet y en *streaming*.
- **FLV:** formato utilizado por el reproductor Adobe Flash para visualizar vídeo en internet, lo que permite que pueda ser visualizado por diferentes sistemas operativos y navegadores web. Se utiliza en la mayoría de repositorios de vídeo como Youtube, Google Video, etc y también admite *streaming*.
- **MKV:** formato de vídeo extensible que actualmente es de los más utilizados para vídeos de alta definición.

Todos estos formatos de vídeo, que se encuentran entre los más comunes, vienen mejor detallados la referencia [15] anteriormente mencionada en este mismo apartado.

Para poder verificar esta nueva funcionalidad del programa, se comprobó que el caracterizador de vídeo analizaba de manera correcta un mismo vídeo, “Casablanca”, transcodificada a los formatos anteriormente comentados. Una vez comprobado que es capaz de reconocer y analizar todos y cada uno de los formatos mencionados, se puede decir que se ha completado con éxito el desarrollo de esta nueva mejora no funcional.

4.1.3 MODIFICACIÓN DEL FICHERO DE SALIDA

Debido a que algunas de las principales aplicaciones para las que puede ser utilizado este caracterizador de vídeo están relacionadas con el mundo Web, ya sea mediante aplicaciones de recomendación de contenidos, plataformas de análisis fílmico, portales de vídeo bajo demanda o gestores de contenido en general, se planteó la necesidad, para garantizar la interoperabilidad con diversos sistemas tecnológicos, de que se proporcionara un fichero de salida en formato XML³.

La modificación del fichero de salida vídeo.(formato), que se generaba en formato .txt, para que se genere en formato .xml, el cual permite una mejor identificación de los descriptores de imagen, de ritmo narrativo y de movimiento del vídeo analizado gracias a que cada uno de los valores que se imprimen en este fichero va asociado a etiquetas que lo identifican, tal y como se puede observar en la *Figura 4.1*.

En la *Figura 4.1*, además de poder observar que cada etiqueta va asociada a cada uno de los descriptores de vídeo, se observa que dichas etiquetas siguen el orden tal y como viene definido en el fichero vídeo.(formato)_descriptor, para facilitar la identificación de cada uno de los valores a nivel usuario.

³ El lenguaje XML es un lenguaje de marcas y utilizado para el almacenamiento de datos en forma legible. Permite el soporte a bases de datos, editores de texto, hojas de cálculo, etc. siendo útil para varias aplicaciones y para su uso en internet.

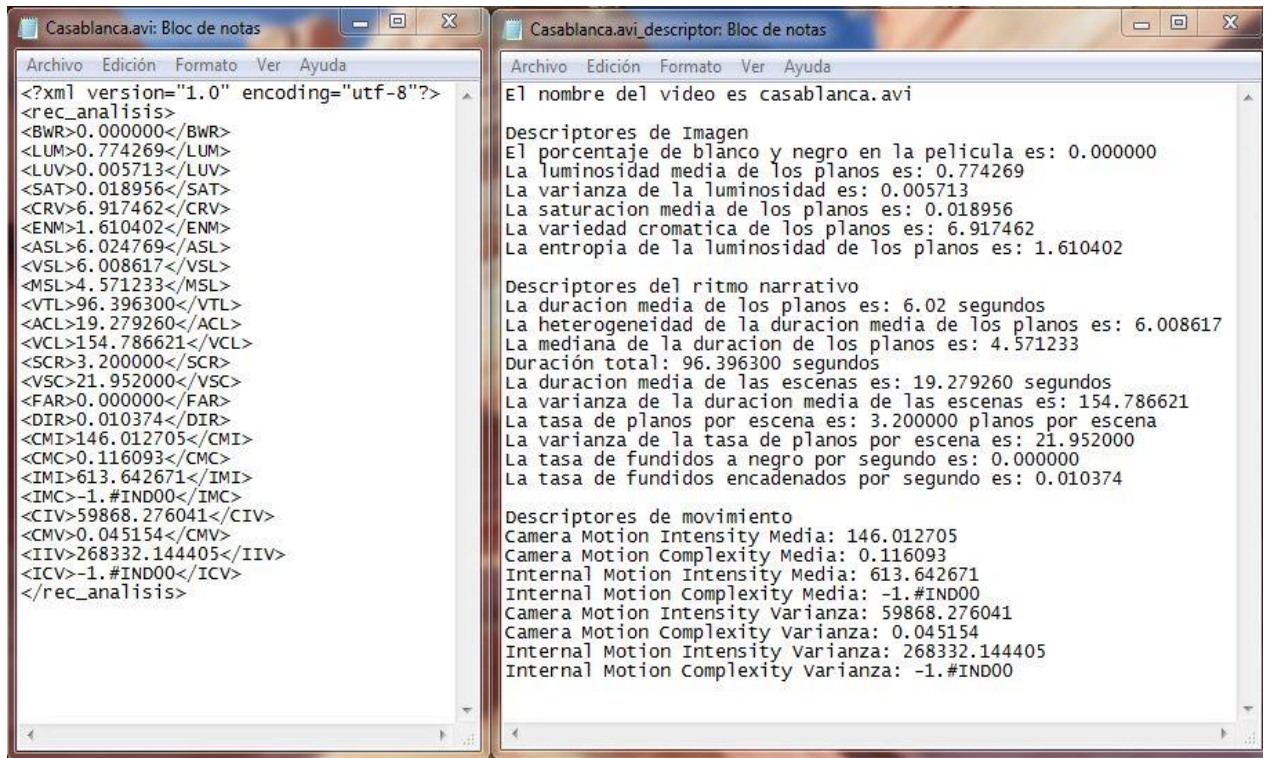


Figura 4.1 Comparativa entre el fichero vídeo.(formato) en XML y el fichero vídeo.(formato)_descriptor

Esta mejora no funcional supone un gran avance para el caracterizador de vídeo ya que se está facilitando, de manera más directa, su utilización en programas web y/o aplicaciones diversas.

4.2 MEJORAS FUNCIONALES

En este apartado del documento se detallan las mejoras funcionales introducidas en el caracterizador de vídeo para mejorar la eficiencia en la segmentación temporal de aquellos vídeos que presentan un mayor número de errores.

Estas mejoras funcionales se diseñaron a raíz de las conclusiones obtenidas en la fase de análisis que corresponde al capítulo 3 de este documento y en el cual se detalla brevemente las razones que llevaron al diseño de estas mejoras. Antes de comentar cada una de estas mejoras, se van a explicar los recursos que se han empleado para poder implementarlas.

Como ya se ha comentado en el anterior apartado, una vez realizada la migración del código, se pueden incluir trazas en el mismo que permitan obtener valores, imprimidos por la consola de Windows, que faciliten la comprensión de los algoritmos y la búsqueda de posibles errores en la detección de cada uno de los cambios de plano.

A continuación se exponen los diferentes tipos de cambios de plano que reconoce el caracterizador de vídeo y la metodología empleada para la depuración de los mismos que se desarrolla con mayor detalle dentro de la implementación de cada una de las mejoras funcionales.

Clasificación de la segmentación temporal

Cada cambio de plano tiene unas características particulares y el código original [1] los detectaba de la siguiente manera:

- **Corte de plano:** había dos formas de detectarlo, *Figura 4.2*, que se correspondían con dos técnicas del proceso de cuantificación de detección de cortes comentado en el estado del arte del capítulo 2, y con la selección de un umbral fijo en el proceso de filtrado:
 1. *Por el movimiento:* se calcula la media armónica del movimiento de una serie de puntos estratégicamente colocados en cada imagen, y se compara el valor entre cuadros consecutivos. Es decir, realmente esta técnica comparaba cada punto de dos cuadros consecutivos, por lo cual era la correspondiente a la comparación del movimiento. Si este valor es superior a un umbral preestablecido, durante una ventana temporal correspondiente a 30 planos, se selecciona la imagen como candidata a preceder un corte de plano. Para comprobar que no se trata de un falso positivo, se calcula la correlación entre planos dos a dos, tomando una imagen anterior y otra posterior al supuesto corte. Si dicha correlación es menor al 50%, es decir, que la similitud entre planos sea como máximo el 50%, entonces se considera corte de plano.

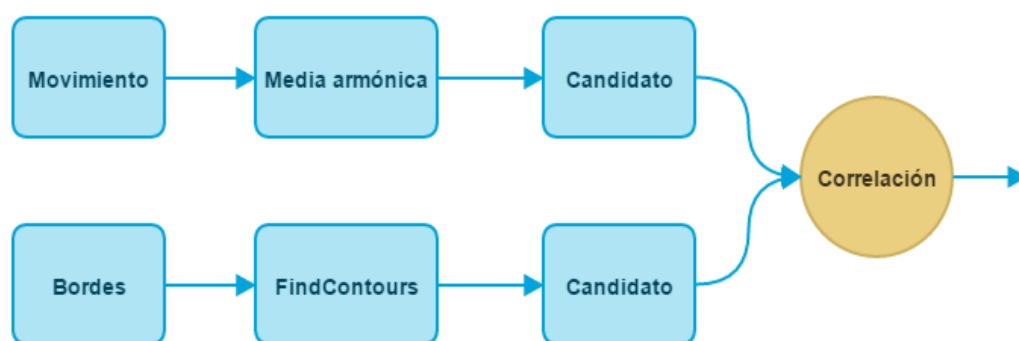


Figura 4.2 Métodos de detección de los cortes de plano en el programa original

2. *Por el número de bordes:* que es la técnica de detección por comparación de bordes ya explicada en el estado del arte, en la cual se calcula el número de bordes que hay en cada cuadro, mediante el comando de OpenCV *FindContours*, basado en el algoritmo de Suzuki [17] y se calcula la diferencia del número de bordes entre dos cuadros

consecutivos. Si esta diferencia es superior a un umbral preestablecido durante una ventana temporal correspondiente a 30 planos se calcula la correlación entre imágenes dos a dos de manera análoga al caso anterior.

- **Fundidos:** como existen dos tipos de fundidos, los fundidos de entrada y los fundidos de salida, se comprueba que se ha producido al menos uno de estos dos fenómenos:
 1. *Fundido de entrada:* es la degradación de la imagen hasta que se convierte en un plano monocolor. En el programa original se detecta comprobando que el número de bordes va descendiendo en cinco cuadros consecutivos hasta quedarse en un cuadro de color negro.
 2. *Fundido de salida:* es el efecto contrario a los fundidos de entrada, un plano monocolor que se degrada progresivamente en una imagen nítida. Se detecta comprobando existe una imagen de color negro y que contando que el número de bordes va aumentando durante cinco cuadros consecutivos.



Figura 4.3 Fundido de entrada y fundido de salida

- **Encadenados:** el programa original trata de detectarlos partiendo del mismo método por el cual detecta los fundidos. Calcula el número de bordes de cada imagen y comprueba vayan descendiendo y/o aumentando progresivamente durante 25 cuadros. Es decir, comprueba que un plano se vaya fundiendo en otro. Tal y como se ha visto anteriormente, este tipo de cambio de plano es el peor analizado por el caracterizador de vídeo.
- **Escenas:** se detectan comparando los *key-frames* evaluando estos como posibles candidatos a cambio de escena calculando las distancias entre sus histogramas. No se ha evaluado con mayor detalle ya que incorpora características semánticas que podrían ser evaluadas para un trabajo futuro.

Metodología de depuración

Una vez vistos los diferentes cambios de plano y cómo los detecta el caracterizador de vídeo original, se realiza un análisis de todos los vídeos incluyendo

trazas en el código para detectar los valores de las medias armónicas, la correlación y el valor de gradiente, correspondiente al valor que se obtiene del método de comparación de bordes. Poco a poco se van imprimiendo por la consola de Windows estos valores y se van añadiendo a las tablas que ya se habían creado en el análisis del capítulo 3, tal y como se puede observar en la *Tabla 4.1*.

Casablanca	Cambios de plano detectados	Comentarios	Corte por media armónica	Correlación	Valor relativo al umbral	Corte por FindCon-tours
29,97 F/s	1. 80 ---> 0:0:2		X	0,338163	154757,29	
	2. 298 ---> 0:0:9		X	0,489689	58319,04	
	3. 1076 ---> 0:0:35		X	0,251568	177841,89	
	4. 1146 ---> 0:0:38		X	0,436585	101096,58	
	5. 1241 ---> 0:0:41	Zoom de cámara, falso positivo			corte por transición	
	6. 1301 ---> 0:0:43		X	0,296388	71616,82	
	7. 1480 ---> 0:0:49		X	0,192134	89249,04	
	8. 1584 ---> 0:0:52		X	0,176886	140693,79	
	9. 1644 ---> 0:0:54		X	0,165915	84671,23	
	10. 1811 ---> 0:1:0		X	0,113688	99387,58	
	11. 2003 ---> 0:1:6		X	0,112825	83860,01	
	12. 2172 ---> 0:1:12		X	-0,030971	136567,55	
	13. 2614 ---> 0:1:27		X	0,060312	158925,11	
	14. 2696 ---> 0:1:29		X	0,036842	179918,13	
	15. 2833 ---> 0:1:34		X	0,066508	177244	

Tabla 4.1 Detección de valores específicos de "Casablanca"

En esta tabla se puede observar que se han añadido 4 columnas adicionales a las creadas en el anterior capítulo. Dichas columnas son:

- **Corte por media armónica:** en esta columna se indica, mediante una X, si el cambio de plano detectado como corte se ha producido por el método de la media armónica.
- **Correlación:** indica el valor de la correlación con el cual el corte de plano ha sido detectado como positivo.
- **Valor relativo al umbral:** si el corte se ha producido por media armónica, indica el valor de dicha media con el cual el corte de plano ha cumplido la primera condición, es decir, ha sido seleccionado como candidato antes de pasar el filtro de la correlación. Si el corte se ha producido por el método de comparación de bordes, da el valor del gradiente con el cual el corte de plano ha sido seleccionado como candidato. En el caso de que no se haya detectado un

corte de plano, se indica como un comentario si se ha detectado un encadenado o un fundido.

- **Corte por FindContours:** en esta columna se indica, mediante una X, si el cambio de plano detectado como corte de plano se ha producido por el método de comparación de bordes.

Como se puede observar, esta tabla es un ejemplo de la tabla creada para el vídeo de “Casablanca” en el cual se tiene todos los cortes de plano analizados y, lo más importante, se observa que el cambio de plano detectado como falso positivo ha sido detectado por el programa por el método de detección de encadenados.

Se ha analizado de la misma manera todos los vídeos seleccionados, de los cuales ya se habían creado las primeras tablas en el capítulo anterior. Como ya se había comentado con anterioridad, los vídeos en blanco y negro son los que mayor valor de la métrica precisión y exhaustividad tienen; por esta razón este Trabajo se va a centrar en los vídeos en color, intentando no perjudicar los valores obtenidos inicialmente de esta métrica para estos vídeos en blanco y negro, para lo cual se efectúan las correspondientes pruebas de regresión.

Contando con el factor tiempo del que dispone este Trabajo y siguiendo los objetivos prefijados al mismo, que se enumeran en el capítulo uno, se decide realizar tres mejoras funcionales, tal y como se dijo en el capítulo 3, en las conclusiones de la fase de análisis: la corrección del error para vídeos de un solo plano mediante la mejora del módulo de generación de descriptores finales, la detección de fundidos en cualquier color y la modificación de la detección de cortes de plano para escenas con mucho movimiento de cámara y montaje rápido.

4.2.1 MEJORA DEL MÓDULO DE GENERACIÓN DE DESCRIPTORES FINALES

Existen diversos vídeos que, por su origen publicitario, de análisis o sencillamente porque el autor así lo ha decidido, se componen de un único plano, es decir, que no tienen ningún cambio de plano.

Una de los problemas encontrados en la fase de análisis fue que, para dichos vídeos, el programa daba un error y se bloqueaba.

Estudiando el código del programa se descubrió que existía un error en la obtención de los valores medios que se pueden ver en los descriptores finales de los vídeos, ya que para su obtención se dividía entre el número de cambios de plano detectados. El problema residía en que cuando intentaba analizar un vídeo con un solo plano, en el mismo existe un único *key-frame* y al hacer la media de los valores se dividía entre cero, por lo cual el caracterizador no era capaz de obtener los descriptores finales y se bloqueaba el programa.

Para solventar ese problema, se modificó el código del programa de forma que para la obtención de estos descriptores finales, en el caso de los vídeos que no tuvieran ningún cambio de plano, tomase como referencia para calcular los descriptores, los valores de ese *key-frame* en lugar de calcular la media. De esta forma el programa ya no se bloquea y permite analizar este tipo de vídeos sin ningún problema.

4.2.2 FUNDIDOS EN CUALQUIER COLOR

Una de las funcionalidades más atractivas y en la cual no se había centrado el caracterizador original era la detección de fundidos en cualquier color ya que, tal y como se ha visto con anterioridad, el programa inicial realizaba la detección de fundidos siempre que fuesen fundidos en negro, que son los más frecuentes en la comunicación audiovisual.

Existen películas que utilizan fundidos en un color específico para provocar sentimientos diversos en el espectador. Si bien no es una técnica muy utilizada en la mayoría de las películas, existen algunos géneros, como el *giallo*, y directores, como Dario Argento, que la utilizan con bastante frecuencia con propósitos expresivos, haciendo que las películas que contienen este tipo de transiciones presente una tasa de acierto en la detección de fundidos realmente bajo.

Un ejemplo de película que utiliza esta técnica de fundidos en un color que no sea negro, es “Gritos y Susurros”, que es uno de los vídeos seleccionados para las pruebas de este Trabajo. En concreto, en este vídeo se puede observar que aparecen dos fundidos en rojo, los cuales son perfectos para comprobar que la mejora introducida en el caracterizador de vídeo es capaz de detectarlos.

Para detectar estos fundidos, se analiza el código original para observar de qué manera detecta los fundidos en negro. Además de lo expuesto en el apartado anterior sobre la cuenta del número de bordes y la diferencia de este número entre planos para comprobar si desciende o parte de cero (el cuadro correspondiente al fundido), se comprobaba que la imagen se ponía en negro ya fuera al principio del fundido o al final del mismo, lo cual facilitaba el conocimiento de si el fundido era un fundido de entrada o un fundido de salida.

Para detectar si la imagen contenía uno o varios cuadros en negro, se convertía dicha imagen en blanco y negro, esto es, se creaba una copia de los fotogramas del vídeo en fotogramas en blanco y negro, lo cual nos permitía comprobar que si todos los píxeles estaban en color negro, era porque ese cuadro era un candidato a formar parte de un fundido.

El primer cambio que se realizó sobre el código original, fue el de dejar de pasar como variable los fotogramas en blanco y negro, ya que el propósito de esta mejora es el de captar fundidos en cualquier color. Una vez que se pasa como variable los fotogramas en sus colores originales, mediante el comando `cvGet2D` de OpenCV se captura el color de cada uno de los píxeles que componen el fotograma en escala RGB.

Esto es posible gracias a que este comando crea un *array* de 3 posiciones, correspondientes a los colores rojo, verde y azul, en los cuales guarda el valor escalar del color del píxel seleccionado.

En la mejora, se capturan los colores de todos los píxeles de cada fotograma y se crea una media de las 3 componentes de color que predominan en ese fotograma. Una vez creadas estas medias, se compara cada valor escalar de RGB de cada uno de los píxeles con cada una de las medias y, si el valor se encuentra próximo a la media, se añade el píxel a una variable suma perteneciente al color rojo, verde o azul. Una vez que se ha analizado el fotograma entero, si el número de píxeles en rojo, verde o azul es superior al 95% de los píxeles que componen el fotograma, se declara el fotograma como candidato a fundido. Se puede observar el proceso completo en la *Figura 4.4*.

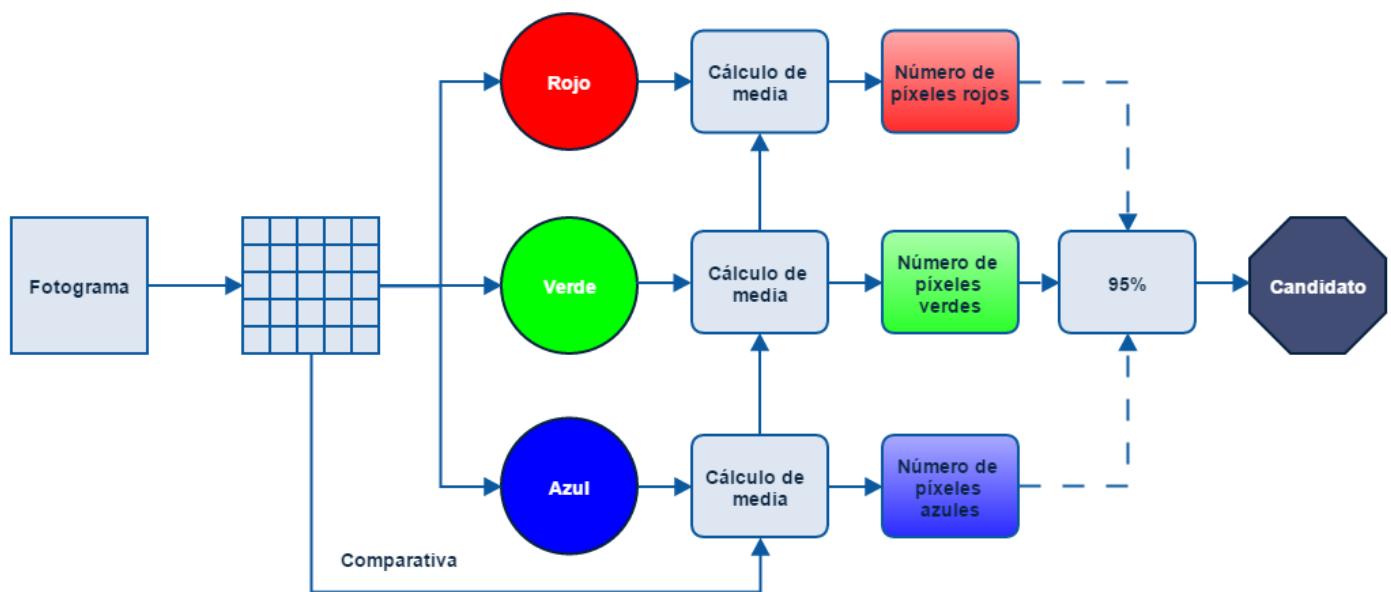


Figura 4.4 Detección de una imagen en un solo color, candidato a fundido

Los siguientes pasos para detectar el fundido se quedan como estaban ya que, al haber realizado estos cambios, siguen siendo válidos para detectar fundidos. Una vez realizada la mejora, se comprueba que el caracterizador de vídeo detecta los fundidos en rojo del vídeo seleccionado. Como el vídeo ya estaba analizado en la fase de análisis, ya se tenía localizado el momento en el que se producía un fundido y el programa no era capaz de detectarlo. Sin embargo, aparte de saber dónde se tiene que producir el fundido, se puede observar por la consola de Windows que cada vez que aparece un fundido se imprime por pantalla *fade in*, *fade out* o ambos.

En concreto, el vídeo analizado de “Gritos y Susurros” contiene dos fundidos en rojo, en el minuto 1 y 46 segundos y en el minuto 9 y 34 segundos. En las tablas obtenidas del análisis no aparecían ambos cambios de plano, sin embargo, una vez vuelto a analizar el mismo vídeo tras esta mejora, se puede observar que actualmente sí que los detecta el programa.

Sin embargo, uno de los contras que aparecieron cuando se consiguió realizar esta mejora en el caracterizador de vídeo, y se comprobó que ya detectaba fundidos en cualquier color, se descubre que ciertas transiciones de vídeo, que no eran cambios de plano, también eran detectadas como fundidos.

La razón por la cual se estaban detectando estos falsos positivos era a que no se había modificado la parte del código en la cual se detectaba el fundido como tal seleccionándolo entre fundido de entrada y fundido de salida. Realizando un análisis sencillo uno de los falsos positivos que habían sido analizados, se comprueba que la razón era que el código original comprobaba en una ventana de 5 cuadros si se producía fundido o no. Por esta razón, cuando había una transición rápida en la cual aparecían, o desaparecían, objetos y por tanto el número de bordes aumentaba o disminuía, esta sucesión de imágenes era analizada como fundido de manera errónea.

La solución fue sencilla de implementar, ya que simplemente con aumentar la ventana temporal en la que se estaba analizando a 20 cuadros en lugar de 5, se descartaban sucesiones rápidas como la que se ha comentado.

Después de implementar esto último, se vuelve a comprobar que el caracterizador de vídeo analiza de manera correcta los fundidos en color rojo y que, además, no introduce falsos positivos al análisis.

Calculando de nuevo las métricas de precisión y exhaustividad, se puede ver en la *Tabla 4.2* cómo han aumentado respecto de los resultados calculados en la fase de análisis.

Análisis anterior		Análisis posterior	
Total cambios de plano detectados	40	Total cambios de plano detectados	42
Detectados correctos	39	Detectados correctos	41
Total cambios de plano correctos	46	Total cambios de plano correctos	46
Precisión	97,50%	Precisión	97,62%
Exhaustividad	84,78%	Exhaustividad	89,13%

Tabla 4.2 Comparativa entre la métrica precisión y exhaustividad de “Gritos y Susurros”

4.2.3 MEJORA DE LA DETECCIÓN DE CORTES

El cambio de plano más utilizado en la edición de vídeo es el corte, es decir el cambio de plano más abrupto que existe y el más común de encontrar en las películas. Por esta razón no es de extrañar que, a pesar de que el caracterizador de vídeo original lo analizaba de manera bastante precisa, se busquen imperfecciones en este análisis para tratar de mejorarlas, especialmente en contextos adversos.

En la fase de análisis de este Trabajo se descubrió que en aquellas películas en las que un personaje o la cámara que grababa la escena se movían a gran velocidad, existían defectos en el análisis de este cambio de plano. En aquellos vídeos en los que el

personaje se movía con gran velocidad existía un número importante de cortes de plano que no eran captados por el programa, como es el ejemplo del vídeo “El Señor de los Anillos”. Sin embargo, aquellas escenas en las que era la cámara la que se movía a gran velocidad, existía un número importante de falsos positivos. Un ejemplo de esto último es el vídeo “The Amazing Spider-Man”, en el cual la cámara sigue al personaje, el cual se mueve a gran velocidad, dando la sensación de que el espectador sigue muy de cerca al personaje en cuestión.

Estos efectos, encontrados en estos vídeos, corresponden con películas bastante actuales que cuentan con un gran número de efectos especiales y que pertenecen al género de acción. Es por ello que para esta mejora, el Trabajo se va a enfocar en el análisis y la comparación de estos dos vídeos.

Tras estudiar el análisis y el estado del arte, se puede comprobar que las técnicas del proceso de cuantificación que emplea el caracterizador de vídeo original son suficientemente precisas y válidas para la finalidad del programa, por lo tanto dichas técnicas no se modifican. Sin embargo, entre las técnicas del proceso de filtrado, el caracterizador usaba el umbral fijo en lugar del umbral variable que permite obtener mejores resultados finales.

Antes de nada, lo primero que se modifica es el umbral fijo inicial para seleccionar los candidatos a cortes de plano por comparación de bordes ya que el valor que estaba establecido en el código original era muy elevado y apenas se seleccionaban candidatos por este método. Posteriormente se modifica la técnica del proceso de filtrado empleada, pasando de umbral fijo a umbral variable.

Analizando los vídeos con mayor profundidad, se puede comprobar que el vídeo de “The Amazing Spider-Man” pertenece a una escena de la película en la cual es de noche y, por tanto, la luminosidad que hay en ella es muy baja. Esto último también se puede comprobar gracias a los descriptores de vídeo, que se obtienen tras el análisis de los mismos, y en los cuales aparece la luminosidad media del vídeo analizado. Por lo tanto, se considera la opción de modificar los umbrales de la media armónica y del gradiente en función de la luminosidad del vídeo.

Otros de los parámetros que se pueden observar en los descriptores de vídeo, son la complejidad e intensidad de la cámara que, además, en estas escenas cuentan con un valor muy alto ya que la cámara que graba la escena se está moviendo a una velocidad considerable. Se considera usar también estas dos variables para realizar el ajuste del umbral para la media armónica y del gradiente.

El otro parámetro que se tiene en cuenta es la intensidad del objeto o personaje que aparece en la escena, ya que en aquellas tomas donde se está moviendo a una gran velocidad, este valor es muy elevado. Este es el caso del vídeo de “El Señor de los Anillos”, donde aparece una persecución entre los personajes y los cortes de plano son casi imperceptibles al ojo humano debido a la cantidad de movimiento que aparece en escena.

Para describir cómo se ha realizado la detección de estos parámetros y el posterior ajuste del umbral, se describe a continuación la obtención de los cuatro parámetros considerados para modificar dichos umbrales:

1. **Luminosidad:** para condicionar el umbral de la media armónica y del gradiente con la luminosidad de cada instante del vídeo, se tiene que calcular la luminosidad de cada cuadro. En el código original, se detectaba la luminosidad de cada *key-frame* y posteriormente se hallaba la media para incluir la luminosidad media en los descriptores del vídeo. Para ello, se crea un nuevo histograma que contiene los datos de cada cuadro dentro del método que detecta los cambios de plano del caracterizador de vídeo. Una vez que se obtienen los datos de cada cuadro/fotograma, se extrae el dato de la luminosidad de cada píxel y se realiza la media con todos los píxeles que componen el cuadro.
2. **Complejidad del movimiento de cámara:** en el código original del caracterizador, se obtiene este parámetro de cada cuadro del vídeo y se guarda en dos histogramas, uno perteneciente a ocho puntos situados en los bordes de la imagen y otro en el interior. El histograma que guarda los valores de los ocho puntos situados en el borde, como se verá a continuación, será de gran utilidad para modificar los umbrales. Además existe un método que devuelve directamente la complejidad de la cámara, pero no va a resultar interesante para esta mejora.
3. **Intensidad del movimiento de cámara:** para obtener este parámetro, al igual que en el caso anterior, existe un método que devuelve directamente el valor de la intensidad de la cámara de cada cuadro respecto al anterior, y que se va a utilizar posteriormente para modificar los umbrales.
4. **Intensidad del movimiento interno:** se obtiene de manera análoga a la intensidad del movimiento de cámara, ya que existe un método directo para obtener este valor de cada cuadro.

Una vez se conoce la manera en la que se obtienen cada uno de estos parámetros condicionantes, se detalla a continuación cómo se utilizan para condicionar los umbrales de la media armónica y del gradiente. Se realizan dos métodos principales, uno de ellos para desechar los cortes de plano que se sabe que son falsos positivos, y otro para detectar aquellos cortes de plano que el programa original no era capaz de percibir.

1. Para desechar falsos positivos:

Se puede observar que la mayoría de los falsos positivos que detecta el caracterizador de vídeo son producidos por un movimiento de la cámara debido a su velocidad y a su cambio de direccionamiento repetitivo en una secuencia del vídeo donde predomina una luminosidad baja.

a) En el caso de la complejidad, se tienen en cuenta los valores que se guardan en el histograma que contiene los datos de los 8 puntos situados en el borde de la imagen, los cuales se actualizan cada cuadro. Dicho histograma es un *array* de 4 posiciones en el que cada posición representa cada cuarto de vuelta de un círculo correspondiente a 360° . Se puede percibir que la cámara está en movimiento ya que los píxeles del borde de la imagen se van desplazando en una dirección que tiene un valor de ángulo clasificado dentro de uno de los cuatro cuadrantes. Como comparar todos los píxeles del borde de la imagen tiene demasiada carga computacional, se escogen 8 puntos que se denominan puntos del marco. Si el programa detecta movimiento de la cámara entre un cuadro y otro, guardará en cada posición del histograma el número de puntos cuyo ángulo pertenece a ese cuadrante. Es decir, como mucho el valor total que se tendría de sumar los valores de cada componente del histograma será de 8, pero puede ser inferior debido a que en alguno de los puntos no se habrá detectado movimiento. En los vídeos donde la complejidad del movimiento de cámara sea baja, los puntos se van a concentrar en una o dos posiciones del histograma mientras que si la complejidad es alta, se distribuirán en tres o cuatro posiciones distintas, como se puede ver en la *Figura 4.5*.

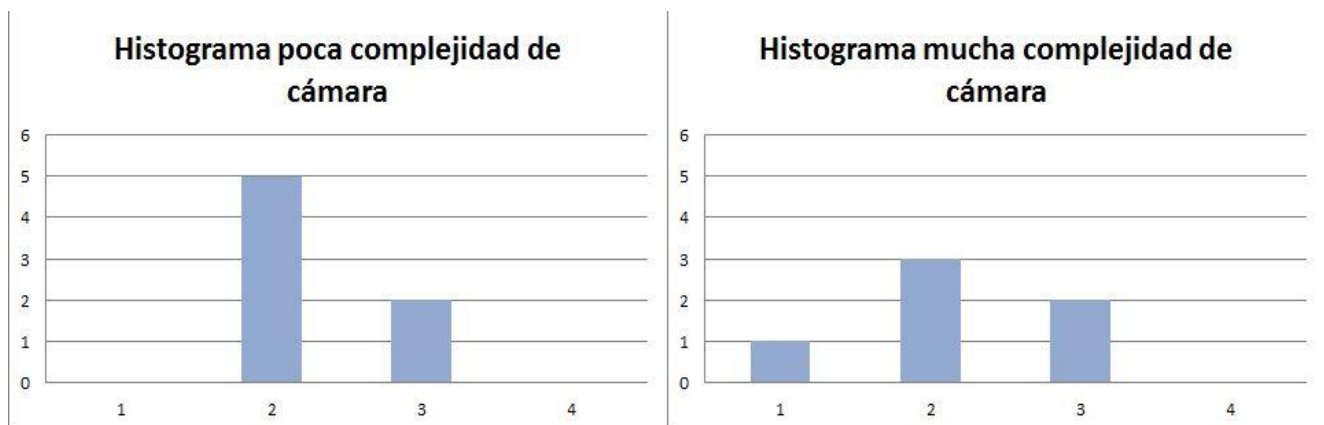


Figura 4.5 Comparativa de histogramas de complejidad de cámara

De este modo, la condición para aumentar el umbral de la media armónica y el del gradiente, y así desechar falsos positivos, será que el histograma de marco contenga al menos tres de sus posiciones distintas a cero.

b) Otra de las opciones que se consideran para desechar falsos positivos es la luminosidad, y es que se puede observar que en el caso de películas con poca luminosidad es complicado diferenciar que la cámara se está moviendo. Por ello, otra de las condiciones para aumentar los umbrales será la de que la luminosidad sea inferior a un valor asignado. Se entiende, una vez realizadas varias pruebas, que con un valor de 0,2 es suficiente para considerar que la luminosidad es muy baja.

c) Finalmente, la última de las características a tener en cuenta es la intensidad del movimiento de cámara. Se considera esta opción ya que no solamente influye el hecho de que la cámara se mueva en varias direcciones, sino que además lo hace con

mucha velocidad. Este parámetro considera, por tanto, que la cámara se esté desplazando rápidamente. Se establece, después de las pruebas realizadas con los vídeos del corpus, que la intensidad de la cámara supere cierto valor, considerando un valor de 500, para aumentar los umbrales.

Si se cumplen estas 3 condiciones, se establece que el umbral que debe superar la media armónica para que el plano sea considerado como corte de plano pase del valor previamente fijado en el código de 4500, a 10.000. El valor de 10.000 se ha establecido observando la gráfica de la *Figura 4.6* que representa el valor de la media armónica de los falsos positivos pertenecientes al vídeo “The Amazing Spider-Man”, que es el vídeo más limitante de los que se han considerado en el corpus.

El valor del umbral de 10.000 es una solución de compromiso para secuencias que cumplen las condiciones excepcionales descritas anteriormente, ya que trata de eliminar el máximo número de falsos positivos sin eliminar cortes de plano correctos que aparecen en otros vídeos y que se pueden ver en las tablas del análisis de los vídeos creados en el proyecto de Excel de este Trabajo.

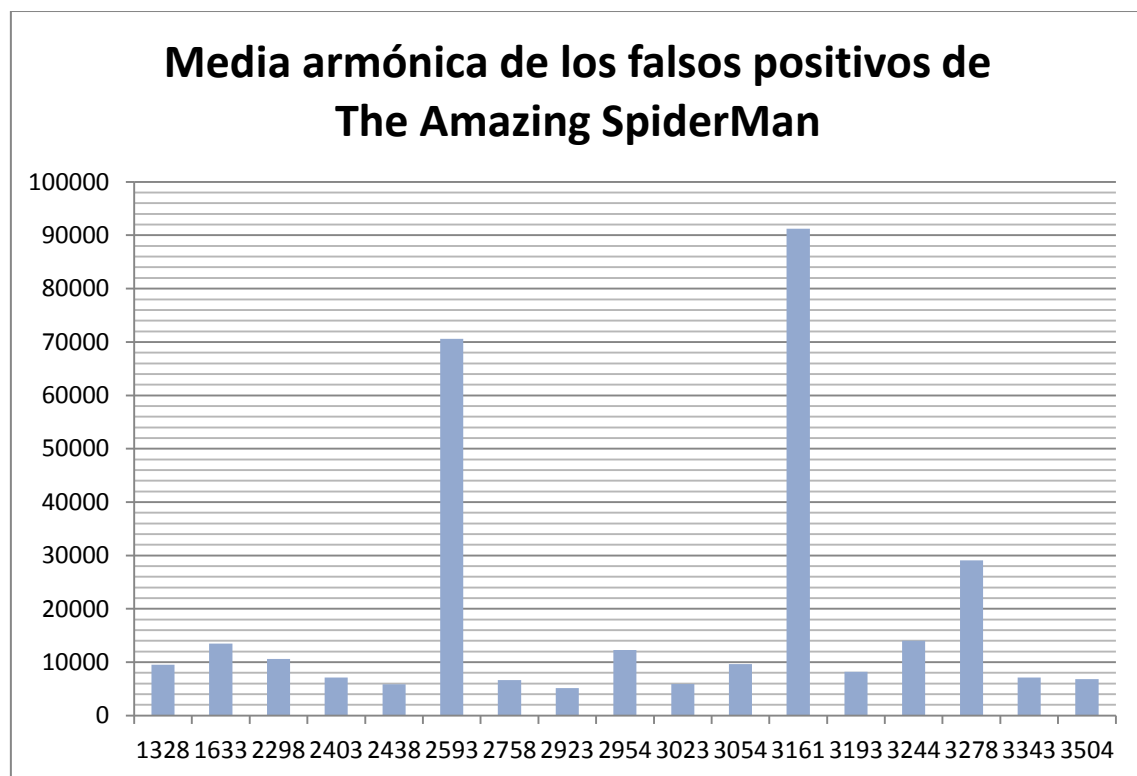


Figura 4.6 Media armónica de los falsos positivos de “The Amazing Spider-Man”

Por otro lado, se establece que el umbral del gradiente para desechar los falsos positivos aumente de 130 a 150 para las condiciones establecidas, ya que con este valor apenas se detectan planos candidatos a corte por el método de comparación de bordes.

2. Detectar cortes de plano no identificados:

La mayoría de los cortes que no detecta el programa original pertenecen a secuencias del vídeo donde aparecen personajes que se encuentran en constante movimiento a gran velocidad y la cámara los está siguiendo. Se diferencia principalmente de los falsos positivos en que en este caso predomina el movimiento de lo que se está grabando frente al movimiento de la cámara que está grabando.

También se puede observar que los vídeos que cumplen estas características, y en los cuales faltan por detectar cortes de plano, cuentan con un valor de la luminosidad bastante más elevado que en los casos donde se detectan falsos positivos. Un ejemplo de estos vídeos es el de “El Señor de los Anillos” donde, como ya se ha comentado anteriormente, aparece una persecución entre los personajes.

Para ello, una de las condiciones a tener en cuenta será que el valor de la luminosidad sea superior a un umbral para modificar el valor del umbral de la media armónica y del gradiente. Se considera que con poner que el valor de la luminosidad sea mayor a 0,2 (valor obtenido experimentalmente), en contraposición con el caso anterior, será suficiente, ya que también se va a contar con más condicionantes.

Uno de esos otros condicionantes es la intensidad del objeto en movimiento. Se tiene que tener en cuenta que estos cortes no detectados suceden, la mayoría, en casos muy extremos donde la intensidad interna, es decir la velocidad con la que se mueven los personajes o los objetos, es muy elevada. Por esta razón se decide que la condición que se debe cumplir para bajar el valor de los umbrales sea que el valor de la intensidad del objeto supere un umbral considerablemente alto. Comparando los valores de esta intensidad entre diferentes vídeos se decide dar un valor de 20.000 como umbral a superar. Uno de los vídeos que se toma como referencia para realizar comparaciones es el de “Titanic” debido a que, como se ha visto anteriormente, su análisis da resultados de 100% en la medida de la métrica precisión y exhaustividad, es decir, presenta unas condiciones óptimas para la detección de cortes.

Por último, se escoge como último condicionante la intensidad del movimiento de la propia cámara. Como ya se ha explicado con anterioridad, la cámara va siguiendo el movimiento de los personajes, aunque no sea la fuente de movimiento predominante. Por esta razón hay que tener en cuenta que el valor de la intensidad de la cámara también debe de ser significativo y así evitar incluir falsos positivos en el análisis. Se considera que debe de cumplir, al menos, las mismas condiciones que para el descarte de falsos positivos explicado anteriormente.

Por lo tanto, si la secuencia de vídeo cumple estas condiciones, se baja el valor de umbral de media armónica, pasando de valer 4.500 a ser de 500, y el umbral del gradiente pasa de 130 a 50.

Finalmente, se decide que también se debería modificar el valor de la correlación⁴, pasando a ser del 50% al 60% o, lo que es lo mismo, que dos cuadros consecutivos se puedan diferenciar hasta en un 60% durante la ventana temporal de 10 cuadros, para ser considerados como cortes de plano.

Una vez implementados estos dos métodos, que constituyen la mejora de la detección de los cortes de plano, cuyo resumen se puede observar en la *Figura 4.7*, se comprueba que ambos métodos mejoran los valores finales de segmentación temporal de los vídeos, quedando implementada la mejora de la detección de cambio de plano. Para ello, se vuelve a calcular la métrica precisión y exhaustividad de los vídeos “The Amazing Spider-Man” y “El Señor de los Anillos” y se comprueba que mejoran, demostrando así que se han eliminado algunos falsos positivos y se han detectado nuevos cortes de plano correctos, *Tabla 4.3* y *Tabla 4.4*.

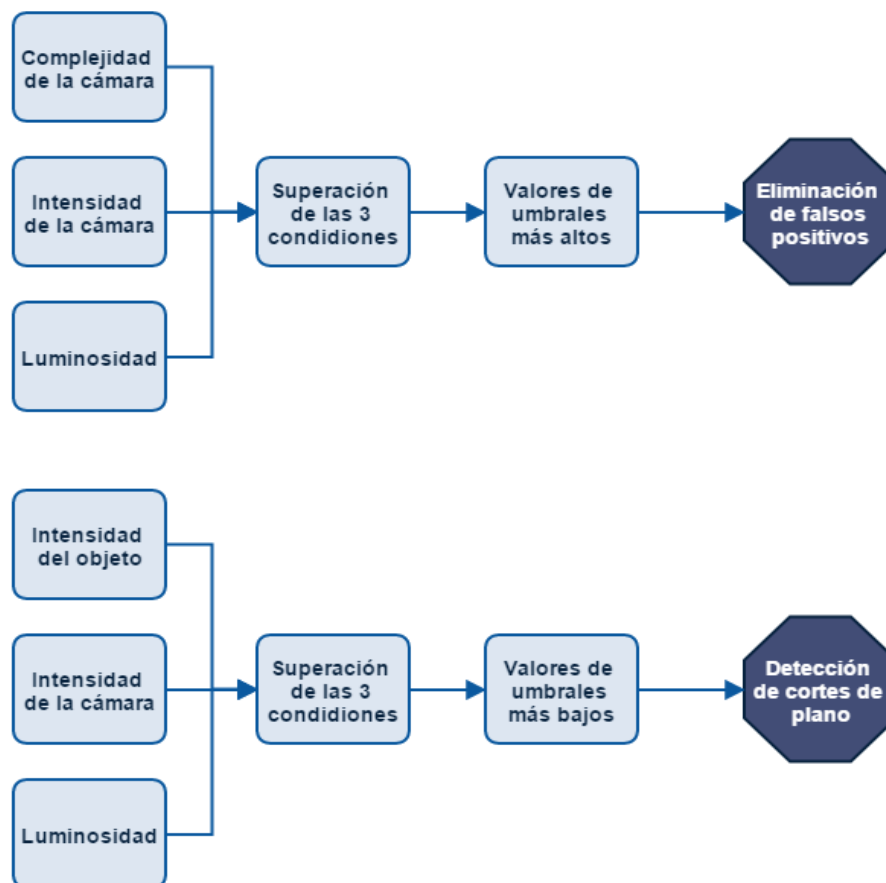


Figura 4.7 Mejora de la detección de corte de plano

En el caso de la *Tabla 4.3*, correspondiente a la eliminación de falsos positivos, la mejora es manifiesta, ya que mejora la precisión de la detección manteniendo los valores de exhaustividad.

⁴ Explicada su funcionalidad dentro del algoritmo de detección de cambios de plano al inicio del apartado 4.2.

Análisis anterior		Análisis posterior	
Total cambios de plano detectados	50	Total cambios de plano detectados	46
Detectados correctos	33	Detectados correctos	33
Total cambios de plano correctos	36	Total cambios de plano correctos	36
Precisión	66,00%	Precisión	71,74%
Exhaustividad	91,67%	Exhaustividad	91,67%

Tabla 4.3 Comparativa entre la métrica precisión y exhaustividad de “The Amazing Spider-Man”

Análisis anterior		Análisis posterior	
Total cambios de plano detectados	84	Total cambios de plano detectados	87
Detectados correctos	80	Detectados correctos	82
Total cambios de plano correctos	96	Total cambios de plano correctos	96
Precisión	95,24%	Precisión	94,25%
Exhaustividad	83,33%	Exhaustividad	85,42%

Tabla 4.4 Comparativa entre la métrica precisión y exhaustividad de “El Señor de los Anillos”

Por su parte, en la *Tabla 4.4*, correspondiente al vídeo de “El Señor de los Anillos” se observa que, a pesar de que el valor correspondiente a la precisión ha disminuido en aproximadamente un 1%, el valor de la exhaustividad aumenta algo más de un 2%, luego en conjunto se asume que la métrica mejora. En un trabajo futuro se podría mirar añadir alguna técnica más experimental del proceso de cuantificación que permita mejor aún más dicha métrica.

4.3 CONCLUSIONES DE LAS MEJORAS

En este apartado se exponen las conclusiones de la implementación de las mejoras en el caracterizador de vídeo.

Haciendo un resumen de las mejoras que se han implementado, estas se pueden clasificar en dos grupos diferenciados, tal y como se puede ver a lo largo de este capítulo:

- **Mejoras no funcionales:**
 1. Migración del código.
 2. Modificación del tipo de vídeo de entrada.
 3. Modificación del tipo de fichero de salida.
- **Mejoras funcionales:**
 1. Mejora del módulo de generación de generación de descriptores finales.
 2. Fundidos en cualquier color.
 3. Mejora de la detección de cortes.

Una vez se ha visto que todas las mejoras introducidas permiten al caracterizador de vídeo mejorar su análisis de los vídeos y le dan un valor añadido al mismo, se analizan de nuevo, al igual que se hizo en la fase de análisis, todos los vídeos seleccionados y se recalcula la métrica precisión y exhaustividad para comprobar que dichas mejoras están aumentando los valores de esta métrica. Para poder observar mejor dicha mejora, se crean las tablas *Tabla 4.5* y *Tabla 4.6* donde se ve una comparativa del antes y después:

Película	C/ByN	Duración	Precisión antes	Precisión después	Exhaustividad antes	Exhaustividad después
Hannah y sus hermanas	C	1:42:21	94,68%	94,68%	98,42%	98,42%
Al final de la escapada	ByN	0:01:25	100,00%	100,00%	84,62%	84,62%
El apartamento	ByN	0:04:29	100,00%	100,00%	28,57%	28,57%
Casablanca	ByN	0:01:36	93,33%	93,33%	100,00%	100,00%
Enemigo público	ByN	0:02:21	100,00%	100,00%	90,48%	90,48%
Con faldas y a lo loco	ByN	0:01:41	100,00%	100,00%	93,75%	93,75%
Apocalipsis Now	C	0:03:08	93,94%	93,94%	86,11%	86,11%
The Amazing Spider-Man	C	0:02:05	66,00%	71,74%	91,67%	91,67%
El Señor de los Anillos	C	0:02:46	95,24%	94,25%	83,33%	85,42%
Titanic	C	0:02:01	100,00%	100,00%	100,00%	100,00%
Gritos y Susurros	C	0:10:03	97,50%	97,67%	84,78%	91,30%

Tabla 4.5 Comparativa de la métrica precisión y exhaustividad entre el antes y el después de los cambios de plano totales

Película	Cortes de plano antes		Cortes de plano después		Fundidos antes		Fundidos después	
	Prec. (%)	Exha. (%)	Prec. (%)	Exha. (%)	Prec. (%)	Exha. (%)	Prec. (%)	Exha. (%)
Hannah y sus hermanas								
Al final de la escapada	96,88	86,11	96,88	86,11	-	-	-	-
El apartamento	100,00	84,62	100,00	84,62	-	-	-	-
Casablanca	100,00	25,00	100,00	25,00	-	-	-	-
Enemigo público	100,00	100,00	100,00	100,00	-	-	-	-
Con faldas y a lo loco	100,00	100,00	100,00	100,00	-	-	-	-
Apocalipsis Now	100,00	93,75	100,00	93,75	-	-	-	-
The Amazing Spider-Man	66,00	91,67	71,74	91,67	-	-	-	-
El Señor de los Anillos	95,18	82,29	94,19	84,38	-	-	-	-
Titanic	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Gritos y Susurros	100,00	95,12	100,00	97,56	0,00	0,00	100,00	100,00

Tabla 4.6 Comparativa de la métrica precisión y exhaustividad entre el antes y después para los cortes de plano y los fundidos

En la *Tabla 4.5* se reflejan los valores de la métrica precisión y exhaustividad totales, es decir, el valor que obtenemos teniendo en cuenta todos los cambios de plano

de los vídeos. En ella se puede observar que la mayoría de los vídeos no alteran su métrica a excepción de 3 de los 5 vídeos que tenemos en color. Teniendo en cuenta que uno de los vídeos en color, “Titanic”, se analizaba con un valor del 100%, el hecho de no haber alterado su métrica se considera como un aspecto positivo. Están marcados con color aquellos vídeos que sí que han cambiado su métrica y se puede ver cómo dicho valor aumenta tras realizar las mejoras en el caracterizador de vídeo.

Tal y como se explicó anteriormente, aunque a simple vista parece que en el vídeo “El Señor de los Anillos” la precisión baja casi un 1%, en exhaustividad aumenta algo más de un 2% lo que hace que en conjunto se considere que la métrica en conjunto aumenta. En el caso de los vídeos “The Amazing Spider-Man” y “Gritos y Susurros” no queda duda de que la métrica ha aumentado considerablemente.

En la *Tabla 4.6* se muestran los valores de la métrica desglosada en las dos mejoras realizadas, tanto para cortes de plano como para fundidos. Se puede ver cómo para el vídeo de “Gritos y Susurros” se pasa de un 0% en la métrica a un 100%, lo cual refleja que actualmente el caracterizador de vídeo detecta fundidos en cualquier color. Por otro lado, tanto para “Gritos y Susurros” como para “The Amazing Spider-Man” y “El Señor de los Anillos”, se comprueba que el valor de la métrica ha aumentado, lo que refleja una mejora en la detección de los cortes de plano.

Debido a que no se ha modificado nada relacionado con los encadenados, no se vuelve a evaluar la eficiencia en la detección de este tipo de cambio de plano, por lo que la referencia para futuros trabajos sigue siendo la expresada en la *Tabla 3.5* del apartado 3.3.

En conjunto, los resultados obtenidos son una muestra de que las mejoras introducidas en el programa original han aumentado la capacidad del caracterizador para analizar vídeos.

5. CONCLUSIONES Y TRABAJO FUTURO

Conclusiones

Este Trabajo de Fin de Grado se ha centrado en el análisis y mejora de un sistema de caracterización estética de vídeos, cuyo fundamento principal radica en técnicas de segmentación temporal de vídeo. A continuación se valora el cumplimiento de los objetivos marcados inicialmente y se exponen las conclusiones desprendidas:

- Se ha analizado el comportamiento del sistema de caracterización de vídeos, realizando pruebas no funcionales y funcionales y obteniendo métricas de desempeño. El resultado del análisis fue satisfactorio, ya que finalmente permitió mejorar los resultados iniciales de precisión y exhaustividad.
- Se han aplicado técnicas de reingeniería, analizando un sistema a partir de su código fuente y obteniendo información de un nivel de abstracción superior, lo cual permitió realizar posteriormente las mejoras necesarias.
- Se han realizado mejoras no funcionales sobre el programa original, mejorando su interoperabilidad:
 1. Se ha migrado el código fuente a la versión más reciente de OpenCV, posibilitando futuras mejoras y garantizando en funcionamiento en las plataformas más recientes.
 2. Se ha aumentado el conjunto de tipos de formatos de vídeo de entrada admitidos.
 3. Se ha incluido la generación de un archivo xml de salida para hacer el sistema más interoperable con otros sistemas de información.
- Se han realizado mejoras funcionales sobre el programa original, mejorando su rendimiento. Estas mejoras han sido evaluadas mediante pruebas cualitativas y cuantitativas:
 1. Se ha corregido el error que imposibilitaba la caracterización de vídeos que constan de un solo plano.
 2. Se ha mejorado la detección de fundidos, incorporando la posibilidad de detectar fundidos en colores diferentes al negro.
 3. Se ha mejorado la detección de cortes, cuyo rendimiento ha mejorado notablemente en escenas con unas determinadas características estéticas. Para ello se han empleado tanto las conclusiones obtenidas del análisis del programa original como las extraídas del estudio del estado del arte.

- Se han realizado pruebas de regresión, garantizando así que las mejoras acometidas no han producido ningún cambio indeseado sobre las funcionalidades iniciales del sistema.

Trabajo futuro

Para poder describir las posibles líneas de trabajo para este programa, se parte de la fase de análisis, en la cual se desglosaron los diferentes tipos de cambios de plano existentes. Se puede observar que las mejoras han ido encaminadas únicamente a los cortes de plano y a los fundidos, quedando pendiente como trabajo futuro la mejora de la detección de los encadenados y las escenas.

- **Encadenados:** al analizar con mayor precisión este tipo de cambio de plano, se observó que era la causa de muchos de los falsos positivos que se producían en el análisis de los vídeos. No solo eso, también se puede ver que en muchas de las películas donde se debería haber detectado encadenados, no se detectaban. Teniendo como referencia la Tabla 3.5 se puede comprobar estos dos fenómenos. Además se puede observar que sí existen casos en los que se detectan de manera correcta algunos encadenados. Por estas razones, es una línea de trabajo futuro muy atractiva ya que se conseguiría aumentar las métricas de precisión y exhaustividad de manera notable debido a que los vídeos que más mejorarían son aquellos en blanco y negro⁵.
- **Escenas:** en este Trabajo se ha comentado muy por encima lo que implica un cambio de escena y ni siquiera se ha llegado a analizar dentro de los vídeos. Sin embargo, si se observa de manera más detenida, se puede comprobar que el caracterizador de vídeo no selecciona de manera precisa la mayoría de las escenas. Esto es debido a que para realizar dicha selección, solo tiene en cuenta los *key-frames* de los vídeos, dejando al margen cualquier otro tramo de los mismos. Una opción muy interesante sería la de condicionar la selección de las escenas según cada tramo temporal de vídeo y no únicamente por los *key-frames* o el resto de cambios de plano que detecta. Sin embargo, una mejora que supondría un importante paso adelante sería la introducción de características de tipo semántico para seleccionar qué planos pertenecen a una u otra escena.

Estas son las dos grandes líneas de trabajo que supondrían una mejora más sustancial del sistema, pero existe una innumerable cantidad de posibilidades en este campo de trabajo, tanto relacionado con la segmentación temporal de vídeos como con la caracterización estética en general.

⁵ Teniendo en cuenta los vídeos escogidos en este Trabajo. De todos modos, de forma general se puede afirmar que el encadenado es una transición especialmente utilizada dentro del cine clásico. Aun así, también sería relevante la mejora en algunos vídeos más recientes que siguen haciendo uso de este efecto.

6. BIBLIOGRAFÍA

- [1] Borja Antona Gómez, *Desarrollo de un sistema automático de anotación y recomendación de vídeos a partir de características de bajo nivel*. ETSIT-UPM, 2003.
- [2] <http://opencv.org/about.html>
- [3] <http://www.microsoft.com/es-es/download/details.aspx?id=30678>
- [4] Rainer Lienhart, *Reliable Transition Detection in Videos: A Survey and Practitioner's Guide*. MRL, Intel Coporation, 2200 Mission College Blvd. 2001.
- [5] Irena Koprinska and Sergio Carrato, *Temporal video segmentation: A survey*. Signal Processing: Image Communication 16, 477-500, 2001
- [6] Y.-N. Li, Z.-M. Lu and X.-M. Niu, *Fast video shot boundary detection framework employing pre-processing techniques*. IEEE Image Processing IET, Vol. 3, issue 3, June 2009.
- [7] J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin and B. Zhang, *A formal Study of Shot Boundary Detection*. IEEE Transactions on Circuits and Systems for Video Technology, VOL. 17, NO. 2, February 2007.
- [8] Raquel Gómez Díaz. *La evaluación en recuperación de la información en línea*. "Hipertext.net", núm. 1, 2003.
- [9] <http://code.opencv.org/projects/opencv/wiki/ChangeLog>
- [10] <http://www.atube.me/video/>
- [11] Raquel Gómez Díaz. *La evaluación en recuperación de la información en línea*. "Hipertext.net", núm. 1, 2003.
- [12] Salton, G. y M. J. McGill. *Introduction to Modern Information Retrieval*. New York: McGraw Hill, 1983.
- [13] Kent A. Et al., *Machine literature searching*. VIII. Operational Criteria for Designing Information Retrieval Systems American Documentation, Vol. 6, issue 2, p. 93-101, Abril 1955.
- [14] <http://academic.csuohio.edu/kneuendorf/frames/editing/Bordwell02.pdf>
- [15] <http://www.ite.educacion.es/formacion/materiales/107/cd/video/video0102.html>

- [16] <https://www.ffmpeg.org/ffmpeg-codecs.html>
- [17] Suzuki, S. and Abe, K., Topological Structural Analysis of Digitized Binary Images by Border Following. CVGIP 30 1, pp 32-46 (1985)
- [18] OpenCV, *The OpenCV Reference Manual* Release 2.4.9, 2014.
- [19] OpenCV, *The OpenCV User Guide* Release 2.4.9, 2014.
- [20] OpenCV, *The OpenCV Tutorials* Release 2.4.9, 2014